ROBUST POWER SPECTRAL ESTIMATION WITH APPLICATION TO ELECTROENCEPHALOGRAPHY IN DISORDERS OF CONSCIOUSNESS

A dissertation

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Tamar Melman

August 2018

ROBUST POWER SPECTRAL ESTIMATION WITH APPLICATION TO ELECTROENCEPHALOGRAPHY IN DISORDERS OF CONSCIOUSNESS

Tamar Melman, Ph.D.

Cornell University 2018

The electroencephalogram (EEG) is a widely-used assay of neural function in research and medicine, whose advantages include high temporal resolution , portability, noninvasiveness, low cost, and ease of use. However, typical EEG recordings contain significant artifact from non-neural sources. These are often of high amplitude and at frequencies that overlap with neural signals. Most EEG analysis therefore involves removing artifacts based on their temporal or spatial features prior to quantitative analysis. While effective, these strategies can remove significant signal along with noise, and can be time-consuming and introduce bias. Here, I develop an alternative approach: tools that are tolerant of these outliers for a mainstay of EEG analysis; namely, spectral estimation. The basic strategy is to apply quantile statistics to multitaper spectral calculation. I then develop confidence intervals for these robust spectral estimates, as well as a novel spectral comparison test. Using simulated EEG data as well as healthy control human recordings, I show that the robust power spectral estimator is less sensitive to artifacts that affect the EEG power spectrum, compared to the standard method. Additionally, the robust approach to spectral comparison resulted in fewer false positives, false negatives, and errors of sign, compared with the standard two-group test approach. Lastly, I applied the robust method to spectrogram data from patients with disorders of consciousness

who exhibit paradoxical activation in response to zolpidem treatment. The analysis shows that the robust method reduces movement and other artifacts even when all the data are used together, while reducing analysis time. I also corroborate findings from previous papers, including the restoration of posterior alpha rhythm seen in one patient after treatment. The main drawback of the robust approach, which is also illustrated by this application, is that it may discard spectral features that are intermittent, especially when they co-occur with artifacts. Thus, the new approach, while it reduces sensitivity to artifact, may be best used as an adjunct to artifact removal, rather than a replacement. In summary, I have created a toolkit that combines robust statistics with spectral methods and demonstrated its utility for EEG analysis.

BIOGRAPHICAL SKETCH

Tamar Melman grew up in the Boston area in a family with a science-heavy background. She undertook a project in her father's physics laboratory the summer before college, coding a statistical analysis toolkit from scratch to analyze optical data. Close ties with research from a young age encouraged her to pursue a bachelors degree in computational biology from Carnegie Mellon University in 2010.

While at Carnegie Mellon, Tamar undertook several research projects. One summer she undertook a project to understand paradoxical ion channel activity in epilepsy using differential equation models of neurons. The results of this project were published in what would become the first of many co-authorships. Another summer was spent at Brown helping to build a database of cis-regulatory genetic networks. She then spent a semester back at Carnegie Mellon working on a protein subunit evolution project, for which she received research honors. She also received academic honors upon graduation.

After obtaining her degree, Tamar joined Dr. Lew Cantley's systems biology lab at Harvard University, where she developed statistical analysis pipelines for high-dimensional metabolomics data alongside development of wet lab skills. With a breadth of experiences, Tamar pursued a PhD in computational biology in 2012 in the Tri-Institute, seeking a project in neuroscience. The excitement of the summer project working on paradoxical neurological behaviors due to nonlinearity of brain systems, as well as Tamar's experience with applied statistics, led her to pursue the current

project. In the lab of Dr. Jonathan Victor, Tamar was able to explore the use of applied statistics to understand the complexity of the human brain.

Tamar was also an active member of both her undergraduate and graduate outdoor clubs. She looks forward to continuing these activities alongside future research ambitions.

ACKNOWLEDGMENTS

Dr. Jonathan Victor, I am forever grateful for your mentorship and your patience. Your commitment to rigorousness and precision in ideas and execution inspire me in my continued pursuit of scientific knowledge. I am fortunate to have gained experience under your guidance.

Dr. Nicholas Schiff, your knowledge and expertise in regards to clinical research have taught me so much. Thank you for all of your guidance, and for bringing your passion to every meeting. I am truly lucky to have been able to learn from you.

To my other committee members, Dr. David Christini and Dr. James Booth: thank you for your essential feedback, guidance, and valuable insight into this project throughout its development. I especially want to express gratitude that you've always encouraged me on my academic journey; I could not have imagined continuing onward without your support. Thank you also to Dr. Uri Eden for your time and effort as an external committee member.

To my labmates: Thank you Dr. Mary Conte for offering guidance and expertise in clinical EEG analysis, and for your ongoing encouragement all these years. Dr. Jon Drover, thank you for your valuable feedback on my first paper submission, and for your ongoing help during my time in the lab. Dr. Andy Goldfine, thank you for all your invaluable contributions to this lab, and for your guidance during various stages of this project. Drs. Keith Purpura, Jon Baker, Sebastian Boie,

Sudhin Shah: it's been a joy working with you all; thank you for providing guidance, advice, and camaraderie. Wanda Hyde, thank you for your helpfulness and kindness; lab wouldn't be the same without you.

To my fellow graduate students: Dr. Tanya Nauvel, you were the first face of the lab I saw when I interviewed for the program, and you've been there throughout. You have been an invaluable resource. Dr. Brian Fidali, thank you for providing data, inspiration, and morale boosts. Zoe Adams, thank you for your help in particular with the Zolpidem data. Jackie Gottschall and Yenchu Lin, thank you for sharing your passion and inspiration; it's been a purrfectly unique experience. You all exemplify the essential nature of quality lateral connections.

To the Tri-Institutional Program in Computational Biology: thank you to Margie Hinonangan-Mendoza, Kathleen Pickering, and Francine Espinoza for your tireless work on behalf of the students.

To my friends throughout my time here, you know who you are and you know how essential you've been. To the outdoor club, and to the office of diversity: thank you for helping me build community during my time here.

And last but not least, thank you a million to my family. Ima and Aba, your unconditional support through all my endeavors has been indispensable. Thank you Ben and Jon for being awesome siblings and mentors. And thank you to Daniel, Emily, Mia, and Shani for giving me extra reasons to smile.

Special thanks goes to the following funding sources for making this project possible: the James S. McDonnell foundation, the National Institutes of Health, and the Tri-Institutional Training program in Computational Biology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

EEG - electroencephalogram

DOC - disorder of consciousness

TBI - traumatic brain injury

UWS - unresponsive wakefulness state

VS - vegetative state

MCS - minimally conscious state

LIS - locked-in syndrome

CRS-R - JFK coma recovery scale-revised

FDG-PET - ^{14}fluoro-D-glucose positive emission tomography

fMRI - functional magnetic resonance imaging

MEG - magnetoencephalogram

EMG - electromyogram

ICA - Independent Component Analysis

CHAPTER 1: INTRODUCTION

Electroencephalography (EEG) is a technique for noninvasive bedside monitoring of brain activity in clinical settings. However, EEG is highly sensitive to outliers from non-neural sources, which pose a barrier to extracting informative frequency content from a recorded signal. This is especially the case when analyzing baseline EEG, as is the case in some diagnostic paradigms such as detecting awareness in brain injured patients (Bai et al., 2017). Outlier removal techniques based on temporal or spatial signatures of non-neural signal are commonly used; however, these can unintentionally remove valuable signal as well as introduce bias. Here, we develop another method based on non-parametric statistical methods and apply it to the study of EEG in disorders of consciousness. Chapter 1 of this thesis provides relevant background information on severe brain injury, the utility of EEG in the study of disorders of consciousness, frequency analysis of EEG data, and non-parametric statistics.

Disorders of consciousness

A disorder of consciousness (DOC) is a persistent disruption in the brain mechanisms that produce conscious experience. DOCs can occur following an acute event, such as a stroke or a traumatic brain injury (TBI), or from chronic conditions such as brain tumors or neurodegenerative conditions. DOCs resulting from acute events are especially challenging to diagnose. 10 million patients annually sustain a TBI and present with a DOC; one study found that 52% remain moderately to

severely disabled a year after injury (Thornhill et al., 2000). Little progress in diagnostic performance has been made in the last several years: (Childs et al., 1993) found a post-trauma misdiagnosis rate of 48% of patients with UWS/VS; (Andrews et al., 1996) found a rate of 43%; and later on (Schnakers et al., 2009) reported 41%. The rate may be even higher, since many of these studies call for more longitudinal follow-up of these patients, and their re-categorization of these patients was done through behavioral assessment using the JFK coma recovery scale-revised (CRS-R) (Giacino et al., 2014). Though the definition and role of consciousness are areas of philosophical and ethical debate (Fins, 2015; Fischer and Truog, 2017; Napier, 2015) the practical implication of an accurate DOC diagnosis is the ability to select appropriate treatments, track improvement, and assess the efficacy of interventions, for a recovery process that is lengthy and resource-intensive (Humphreys et al., 2013).

Behavioral diagnostic approaches

Several behavioral assessments have been developed by clinicians for DOC diagnosis and prognostic assessment. One of the earliest and most widely used assessments is the Glasgow Coma Scale (GCS), which measures motor response, verbal performance, and eye opening (Teasdale and Jennett, 1974), with a few drawbacks such as its reliance on verbal response (Bordini et al., 2010). However, it continues to have utility in particular during the first hours after patient admission (Bai et al., 2017). The current clinical standard of care for assessing DOCs throughout recovery involves using the CRS-R (Giacino et al., 2014) scale mentioned in the

previous section. This assessment uses behavioral markers in response to environmental stimuli: tracking visual stimuli, auditory responses, etc. The CRS-R is particularly effective for distinguishing between unresponsive wakefulness (UWS) and minimally conscious states (MCS), two states which can be difficult to differentiate but diverge significantly in their prognosis. Whereas UWS implies no preserved activity above the brainstem, MCS exhibits fluctuations in markers of consciousness (Giacino et al., 2004) including periods during which activity appears to be absent, posing a diagnostic challenge to clinicians (Teasdale and Jennett, 1974).

In many cases, behavioral assessments like the CRS-R are informative; however, one of their biggest drawbacks is their reliance on the assumption of intact motor or sensory pathways. Deficiencies in either or both can occur after acute brain injury and can limit a patient's observable responsiveness, even if they retain activity in their cognitive pathways. Adequate patient care relies not only on current consciousness state, but also on potential for future recovery. Additionally, since MCS patients demonstrate fluctuating consciousness, assessments may come to different conclusions at different times (Gosseries et al., 2011; Schnakers et al., 2009). Lastly, assessment is subjective and clinician-dependent, an issue for a diagnosis on which further rehabilitiation options and insurance coverage hinge. This motivates the use of neural assays to complement behavioral assessments.

Neuroimaging diagnostic approaches

Non-invasive in vivo neural assays can be broadly categorized as metabolic or electromagnetic. Metabolic assays such as fluoro-D-glucose positive emission

tomography (FDG-PET) and functional magnetic resonance imaging (fMRI) have shown great promise for DOC diagnosis (Perri et al., 2016). Both indirectly measure the spatial distribution of brain function, and are therefore useful in probing brain activity with high spatial resolution. Electromagnetic assays on the other hand directly probe electrical activity in the cortex; these include the magnetoencephalogram (MEG) and the electroencephalogram (EEG). The MEG measures magnetic activity near the scalp, which correlates with electrical activity in the cortex. It is less sensitive to distortions due to volume conduction; however, it requires a shielded room and expensive equipment.

The EEG is therefore of interest for its portability, low cost, and ease of setup relative to the MEG. In addition to the aforementioned advantages common to electromagnetic assays, EEG has fewer constraints on implementation than FDG-PET or fMRI. FDG-PET requires specific patient conditions: 4 hours of fasting prior to injection, adequate hydration, and a 30 minute uptake period before imaging can be performed (Brown et al., 2014a). fMRI is contraindicated with ferromagnetic implants and pacemakers, and requires expensive equipment and perfect patient stillness. EEGs can be set up immediately on patients with any metabolic status and is compatible with nearly all medical devices. EEGs are also more ubiquitously available at most hospitals and can be transported to the bedside; they can be used on patients whose conditions preclude movement stabilization; and it is relatively easy to set up. These properties of EEG motivate its use for studying disorders of consciousness (Cruse et al., 2011), and to predict clinical outcome after TBI (Bagnato et al., 2015). Specifically, the clinical goal is to find reliable EEG correlates of

consciousness in order to reduce reliance on behavioral assessments for DOC diagnosis. For non-clinical purposes, finding EEG correlates of consciousness gives researchers better tools to study the processes that collectively produce the experience of consciousness.

The electroencephalogram

Biological origin of the EEG

The human brain consists of about 100 billion neurons, connected to each other in interconnected networks that take in, process, and store information from the outside world through highly interconnected network activity. The briefest unit of electrical activity is the action potential, which takes place with a time course of less than 1ms. Individual spikes don't appear in voltage timecourses from the scalp due to this short time course as well as volume conductance effects rendering axonal spiking neutral from a distance, since dipoles created at the leading edge and trailing edge of an action potential cancel out.. Rather, the contribution is from synaptic potentials, which happen on longer timecourses of 10ms or more. Specifically, the potentials recorded in scalp EEG originate from synaptic activity from cortical pyramidal neurons, a population of cells that lie parallel to one another in layer V of the cortex . The aligned apical dendrites and axons in each patch of cortex ΔV, located near the cortical surface, produce extracellular current activity which, when observed from a distance (as on the scalp), can be approximated as a dipole moments. The surface EEG is the integral over all the contributions of dipoles along the cortical surface,

with little to no contribution from non-cortical sources. However, since thalamocortical neurons terminate on pyramidal neurons, as do cortical interneurons, the pyramidal cortical layer integrates information from other structures and distal cortical regions (Sarnthein and Jeanmonod, 2007). This, the pyramidal cells essentially behave as a neural antenna to transmit electrical activity integrated over brain networks through the tissues of the cerebrospinal fluid, through cracks and openings in the skull, and through the skin, onto the surface of the scalp (Kirschstein and Köhling, 2009) (Nunez et al., 2016).

Electrodes can pick up and record the voltage fluctuations either within the brain tissue (LFP), under the scalp (ECoG), or on the surface of the scalp (EEG). The latter is the oldest recording technique: the first human EEG was recorded in 1924 by Hans Berger, using an analog device with a needle that traced the signals onto a sheet of paper (Berger, 1929). The modern EEG consists of a net of electrodes placed on the head which feed into an amplifier and are digitized. Although the device has been modernized and digitized, the fundamentals are largely the same as the original EEG as recorded nearly 100 years ago.

Since the alpha rhythm was first recorded (Berger, 1929), several EEG rhythms have been identified and correlated to neurological processes. The specific rhythms generated by the neural circuitry correspond to different physiological processes in the brain. Different states of arousal and cognition have different characteristic activity patterns that are consistent between human subjects (Schiff et al., 2014) (Forgacs et al., 2017) (St. Louis et al., 2016). There is some inter-subject variability, but general features are common between subjects in the same state. The

EEG also provides low-resolution spatial information about the origin of particular rhythms.

Researchers continue to develop new quantitative EEG analysis approaches for both clinical and research settings. In addition to disorders of consciousness, applications in the clinic include monitoring patients during anaesthesia (Jameson and Sloan, 2006) (Kreuzer, 2017), detecting non-convulsive seizures (Herman et al., 2011), and sleep studies (Campbell, 2009). It is also used in research of other neurological and psychological conditions including neurodegenerative diseases (Sarnthein and Jeanmonod, 2007), dementia (Bonanni et al., 2008) (Brown et al., 2014b), mood disorders (Lee et al., 2018), schizophrenia (Dvey-Aharon et al., 2015), and others. The number of applications for EEG will undoubtedly continue to proliferate as analysis and noise reduction techniques improve. Here I focus on analyses applied to study disorders of consciousness.

Quantitative EEG analysis: frequency decomposition

Since the EEG records voltage at high enough temporal resolution capture brain dynamics that correspond to physiological processes, the frequency content of the signal is a relevant metric for both clinical and basic research. In clinical research into DOCs, the frequency content of the EEG is often used to assess dynamics corresponding to awareness and wakefulness. (Forgacs et al., 2017) demonstrates how specific power spectral features can be used to classify subjects according to preserved consciousness.

There are two main limitations to obtaining accurate spectra on EEG signal. The first is a mathematical and statistical issue arising from the definition of the power spectrum itself: namely, that it assumes a stationary signal of infinite length. The second limitation to obtaining EEG spectra is the signal recorded from the scalp contains signal from many sources, including environmental electrical sources, muscle artifact (EMG), eye blinks, sweat, movement, etc., alongside the neural signal of interest. Two main strategies can address the mathematical and statistical limitations of the frequency transform of a finite nonstationary signal. One involves reducing the limitation of non-stationarity by cutting the signal into short segments over which the signal can be assumed to be locally stationary, and averaging over the transforms of the segments. The other addresses the issue of spectral leakage into neighboring frequency bins that is the result of taking the frequency transform of a finite signal. This can be addressed by applying the multitaper method (Thomson, 1982), where a set of orthogonal windowing functions are applied to each segment: these have the property of minimizing frequency-domain side lobes when averaged together, at the expense of frequency resolution. Here, I focus on the second, non-mathematical limitation of obtaining power spectra from EEG signal: that of non-neural artifact contaminating the recorded EEG.

Noise in the EEG

The biggest drawback to frequency analysis of EEG is that the electrical signals generated by the cortical neurons are small in amplitude relative to artifacts from non-neural sources. Muscle contractions are produced by electrical signals that

then is conducted through to the skin and picked up by the EEG, with a higher contribution from sources proximal to the head. Swallowing, jaw clenching, and even holding one's head upright produce signals picked up by the EEG as a high frequency broadband artifact called the electromyogram (EMG). The subject need not even be very active in order to produce significant artifact (Whitham et al., 2007) (Whitham et al., 2008) Furthermore, eye blinks move the base of the optic nerve, a source of low frequency noise. Sweat can also interfere with the impedances of the electrodes and thereby affect the recorded voltages. Line noise from A/C oscillation of electricity in wires and devices in the room can produce a sharp peak at 60 Hz. The EEG is affected by mechanical movement of the wires. Even in cooperative healthy human test subjects, recordings can be dominated by noise. Notably, most of the noise sources mentioned produce voltage changes much larger than the underlying neural signal.

Different artifacts have different characteristic temporal, , spatial, and frequency-domain behaviors. Line noise, for example, persists across all segments and all spatial components but is contained in a narrow frequency band. EMG produces noise that affects many frequency bands but can be temporally intermittent and mostly affects the off-midline channels. Eye blinks affect few frequency bands, are intermittent, and affect only the frontal channels. Noise that pervades all three domains is typically impossible to remove. For noise that is concentrated in one of the domains, however, removal is possible: for spatially concentrated noise, independent component analysis (ICA) can be used to obtain a set of dominant spatial components, which can be selectively removed leaving a set of relatively clean spatial

components from which to reconstruct the signal. For noise that is concentrated in the temporal domain--artifacts that come on intermittently for a brief time and then stop--the artifact-containing segments can be removed from analysis. For noise that affects a narrow frequency band, filtering can remove known outlier peaks. All of these methods are commonly used; temporal cleaning is especially necessary for even the cleanest awake recording (since avoiding blinking, swallowing, and other small movements is impossible), and is sometimes paired with simultaneousvideo recording of a subject to identify movement-based artifacts.

Although these techniques are effective, they have some drawbacks: by removing parts of the signal, especially temporally or spatially, relevant signal could be lost along with it, and it may introduce bias. In cases of extremely noisy data, artifact removal may eliminate too much real signa to perform meaningful analysis. Because of these limitations, we explored a third approach to handling noise which can be used either in conjunction with or independently of, existing techniques.

Robust statistical estimates

Previous approaches to robust power spectral estimation

Several approaches have been explored to robustly estimate power spectra for geophysical, climactic, and other long-range time series. One approach is outlier weighting, as (Chave et al., 1987) presents, where a weighting function is applied to iteratively downweight outliers that exceed some threshold above the robust average. While this was demonstrated to work on data from time variation of the Earth's

magnetic field, it is not discussed for the limitations of EEG data, and is only shown on datasets with about 150 data segments and about 10-20% outliers (Chave et al., 1987). (Kleiner and Martin, 1979) apply approaches based on pre-whitening and m-estimators to better distinguish narrow-band peaks; this is not typical of EEG, however, and again the number of data segments is far smaller. Additionally, these approaches assume signal stationarity, which is not typical of EEG.

EEG-specific approaches take into account the specific characteristics of the data. In EEG, several techniques have been developed for reducing noise in spectra ; they typically do this by detecting and removing or downweighting outliers based on PCA (Shi et al., 2013), ICA (Radüntz et al., 2015) (Delorme et al., 2007) (Li et al., 2006), thresholding (Krauledat et al., 2007), Bayesian classification (LeVan et al., 2006), and filtering (Proekt, 2018), most commonly . As mentioned, outlier detection and removal is one possible approach and can unintentionally remove data along with noise. Until now, there has been no formalized robust modification of the Thompson (Thomson, 1982) multitaper method--one that can compliment existing approaches. Here, we develop one such robust approach.

The median approach

The power spectrum is the mean power over all included segments of the signal, giving an estimate of the power spectrum of the generative process. Stated more generally, the estimated power spectrum is the location estimate of the powers from the signal segments at each frequency. It is then allowable to use a non-mean location estimate.

As a location estimator, the mean statistic is sensitive to even single large outliers (Box, 1953). More formally, it has a breakdown point of 1/n, where n is the number of samples (Hampel, 1968). Not being constrained to the mean, we chose a more robust estimator. Since the median has a breakdown point of 50%, this led us to refocus the standard analysis methods around the median statistic. A breakdown point of 50% would mean that in practice, up to half of the data segments included in power spectral analysis can contain artifact without affecting the median spectral power estimate, while the mean would give an estimate heavily affected by the artifact. Furthermore, since the estimate of location is applied to each frequency bin individually, the median would treat different types of artifacts affecting different frequency bands separately, allowing for potentially "clean" results even with artifact in more than 50% of the segments -- provided that they occur at different frequencies. Lastly, this approach could be used to rapidly analyze data sets that contain noise, cutting down on preprocessing time significantly; or it could round out the toolkit of data cleaning approaches, by providing a frequency-domain approach that can be combined with temporal and spatial methods of data cleaning.

It is worth mentioning that the median is a specific case of quantile estimator, and the robust approaches laid out in this thesis are applicable to quantile statistics more generally. Specifically, I apply a Bayesian approach to power spectral estimation, compatible with the median and quantile statistics more generally. These include Bayesian confidence intervals on the median spectrum as well as a Bayesian test for comparing two spectra.

Why a Bayesian approach?

The robust approach developed here assumes that EEG data deviate from normality due to the presence of outliers, and therefore, that non-parametric approaches are more informative for EEG. Well-established non-parametric statistics, such as the bootstrap and rank-sum, are frequently used in cases of known data deviation from normality. Here we set out to derive confidence intervals and a two-sample test that have empirical meaning when applied to the median.

Since the median is a statistic derived from the observed data, a Bayesian approach to confidence intervals and hypothesis testing based on the median is the most direct: using the Bayesian approach explicitly defines the observable data as a fixed variable and the median as a distribution (Bernardo, 2011). Using a flat prior, we can produce a distribution-free test by constructing the posterior on the parameter of interest; in this case, the median--or the difference in medians. Additionally, though here an uninformative prior is used, the Bayesian approach allows for the incorporation of other priors, providing additional flexibility.

Some non-parametric tests behave similarly or converge to the Bayesian approaches applied and developed here; in chapter 3, non-parametric test convergence to the Bayesian approaches is shown to be correlated with robustness. Therefore, when using the median statistic, the Bayesian methods are not only practical for use on real data, but serve as a standard against which to test other robust approaches.

A robust time series analysis toolkit

All together, the above methods give us a comprehensive robust toolkit to analyze timeseries data in the presence of intermittent, tailed noise. The toolkit fits into the existing analysis pipeline and with existing power spectral analysis approaches, in order to potentially retain more data from clinical EEG recordings.

The methods developed in chapters 2 and 3 are publicly available as a MATLAB toolkit based off of the Chronux toolbox (Bokil et al., 2010).

Current work on median robust statistics in EEG

The median statistic has been applied in (Wong et al., 2011) on coherence spectra to produce a cleaner coherogram. Though the focus in this thesis is on single-channel spectral measures, this indicates results might be promising for multichannel data as well. Studies sometimes use non-parametric tests between subjects: Lee et al. (Lee et al., 2018) apply the non-parametric rank-sum statistic to estimate the difference in amplitudes of frequency bands across subjects, between depressive and neurotypical cohorts. And following the publication of (Melman and Victor, 2016) (see chapter 2 of this thesis), (Yan et al., 2017) published a visual tool that applies the median across channels to better capture spectrographical signatures of seizures for use in training new neurology residents to identify ictal events using EEG.

These and other avenues for applying median statistics to EEG data are worth exploring; however, in this thesis I focus on developing and applying robust median-based statistics to multitapered power spectral estimation of EEG datasets. The

methods presented here are novel approaches that address problems specific to EEG data.

The rest of the this thesis is organized as follows. Chapter 2 consists of a first-authored paper on a method we developed that applies non-parametric statistics to electroencephalographic power spectra. Chapter 3 consists of a second paper, ready for submission as of the writing of this thesis, on an extension of the principles from the first paper that applies the non-parametric approach to the comparison of two spectra. Chapter 4 demonstrates the application of the robust approach to resting state EEG data from a clinical study. Chapter 5 is a discussion of the findings and prospective applications.

CHAPTER 2: ROBUST ESTIMATION OF POWER SPECTRA IN EEG

This chapter addresses the issue of outliers in EEG data and their effects on the power spectrum. It introduces a new approach to handling these outliers using a method based on robust statistics: namely, the median multitaper method for power spectral estimation. It develops corresponding confidence intervals using a Bayesian approach.

The body of this chapter was published in 2016 in Journal of Neuroscience Methods.

1. Introduction

Electroencephalography (EEG), a technique for recording the electrical activity of the brain via surface electrodes, is a commonly used assay of brain activity in research and clinical settings. Well-recognized advantages of the EEG include its high temporal resolution, noninvasive nature, and ease of use. However, it is also highly sensitive to electrical activity from non-neural sources, such as eye movements, muscle activity, electrode movement, and electric fields from the environment. These sources generate signals that corrupt the underlying neural signal, and are difficult, if not impossible, to avoid.

For many research applications, and increasingly for clinical applications (Schiff et al., 2014), spectral measures are used to analyze EEG characteristics (Mitra and Pesaran, 1999). Since activity in specific frequency bands often has direct biological interpretations, the power spectrum is of particular interest. However, since the raw EEG signal is contaminated by non-neural sources, obtaining reliable estimates of the power spectrum that reflects underlying brain activity is not straightforward.

Computation of the power spectrum typically involves segmenting the continuous signal, applying Fourier analysis to each segment, and calculating the mean power of the components at each frequency. Fourier components arising from segments contaminated by typical artifacts (e.g., muscle and eye movements) are typically large relative to those of segments that only contain the neural signal, and therefore bias the mean upwards. This problem is usually solved by removing these artifacts, by a combination of manual identification of artifact-containing segments and automated means, such as independent component analysis (ICA) (Makeig et al., 1996); however this can be labor- and time-intensive, subjective, or reduce statistical power.

Here we describe an alternative approach to this outlier problem, via the use of robust statistics. Specifically, we focus on the median and other quantile-based statistics. Via simulations and application to real EEG data, we show that this approach can recover the power spectrum of the underlying signal even in the presence of substantial artifact. Finally, we provide code that extends the Chronux

(Bokil et al., 2010), (Mitra and Bokil, 2008) toolbox to carry out these computations, including the calculation of Bayesian confidence intervals.

2. Methods

2.1. Algorithm

2.1.1. Modified multitaper method

A power spectrum is typically estimated from a measured time series by cutting the time series into segments, applying Fourier analysis to these segments, and averaging the power in each frequency bin across segments. The true value of the power spectrum is the limit of this process as the length and number of the data segments tend to infinity. However, in practice these segments are finite in length and limited in number, so power spectral estimates are necessarily biased (resulting from spectral leakage due to the finite length of the data segment) and imprecise (due to the finite number of data segments).

The multitaper method (Prieto et al., 2007), a power-spectral estimator that we use as a starting point for our approach, tackles the tradeoff between this bias and variance in a way that is optimal for Gaussian signals. The method minimizes spectral leakage (the artifactual spreading of power from one frequency bin into its neighbors), by windowing each segment by an orthogonal set of functions, the Slepian tapers. For further background on the multitaper method see (Mitra and Bokil, 2008; Mitra and Pesaran, 1999; Thomson, 1982). Chronux is a freely available MATLAB toolbox that

provides convenient implementations of the multitaper method, which we extend with an implementation of the robust approach.

The standard multitaper method consists of the following steps: (1)multiplying each data segment by each of the tapers, (2) applying Fourier analysis to these products, (3) averaging over tapers within each trial, and (4) averaging over trials. To formalize this, we denote the original signal by $X(t)$, which is then cut into B segments, denoted as $x_1(t), \dots, x_b(t), \dots, x_B(t)$, each of length T. These segments are non-overlapping, but need not be contiguous. We denote the K Slepian tapers by $a_1(t), \dots, a_k(t), \dots, a_K(t)$. (The choice of K is driven by the desired spectral resolution and data length; a common choice for 3-second-long segments, and the Chronux default, is $K = 5$). With this notation, the standard multitaper estimate of $S_x(\omega)$, the true spectral power at frequency ω, is defined as:

$$\hat{S}^{standard}{}_x(\omega) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{T}\left|\int_0^T x_b(t)a_k(t)e^{-i\omega t}dt\right|^2 \quad [1]$$

We denote the power estimate for a single sample b and a single taper by $S_{b,k}(\omega)$:

$$S_{b,k}(\omega) = \frac{1}{T}\left|\int_0^T x_b(t)a_k(t)e^{-i\omega t}dt\right|^2 \quad [2]$$

With this notation, the standard spectral estimate takes the form

$$\hat{S}^{standard}(\omega) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{K}\sum_{k=1}^{K}S_{b,k}(\omega). \quad [3]$$

Thus, the standard multitaper estimate is a nested mean: first a mean over the K tapers within each segment to obtain the estimate $\hat{S}_b(\omega) = mean\left(\{\hat{S}_{b,k}(\omega)\}\right)$, and then a mean over the B segments:

$$\hat{S}^{standard}(\omega) = mean\left(\{\hat{S}_b(\omega)\}\right). \quad [4]$$

Since our goal is to reduce the effect of outlier estimates from each segment, we replace the mean over segments by a robust estimator, resulting in the estimated power spectral quantity $\hat{S}^{robust}(\omega)$. There are many possible choices for the robust estimator – for example: an estimator based on the h^{th} quantile, a trimmed mean, or a Winsorized mean (Huber, 1963). While the present framework applies to all of these choices, estimators based on quantiles are more readily amenable to computation of Bayesian confidence intervals (see below), and we therefore focus on these, both in the illustrations below and in the MATLAB toolbox. We denote the estimator based on the h^{th} quantile as $\hat{S}^{quantile\_h}(\omega)$. Note that $h = 1/2$ corresponds to the median; this is the default value in the code.

Even for Gaussian data, the median power of the tapered estimates does not equal the mean power. This is because spectral estimates are approximately distributed as chi-squared, which is positively skewed. As shown in the Appendix, we can take the skewing into account by dividing the median power by a data-independent scale factor. Furthermore, scale factors can be derived that convert not just the median (0.5 quantile), but any quantile, into mean power. The appendix details the calculation of these scale factors, which is implemented in the MATLAB module `analytical_scalefactor_Robust()`.

Including this scale factor yields our main result, the robust spectral estimate:

$$\hat{S}^{quantile\_h}(\omega) = quantile\_h\left(\{\hat{S}_b(\omega)\}\right)/C(h,d,B), \quad [5]$$

where $C(h,d,B)$ is the scale factor for quantile h; d is the number of degrees of freedom ($d = 2K$ for typical frequencies, $d = K$ for DC and the Nyquist frequency); and B, as above, is the number of segments.

Notably, the quantile is applied to the B power estimates from each segment (replacing the outer operation in Eq. [3]); within segments, the step of computing the mean over the tapers remains unchanged from the original method. There are two reasons for this choice: 1) if artifact is present in a segment b, it is likely to affect many of the tapered estimates from that segment; and 2) the K Slepian tapers were designed to be used together to capture all of the power within a frequency bin. The toolbox supports the alternative strategy of computing the median across all tapered estimates or the "two-tiered" median across tapers and then across trials, but as we see no principled reason for this, it is not the default. The robust approach is also applicable to other spectral estimation methods, such as Welch windowing.

2.1.2. Confidence interval estimation

Standard nonparametric approaches to confidence interval estimation (Thomson, 2007) are based on resampling strategies, such as the jackknife or the bootstrap. These approaches are appropriate for the mean, which depends smoothly on the data – a necessary condition for the jackknife or bootstrap to be valid.

However, since quantile-based estimates do not depend smoothly on the data, an alternative approach is needed.

Our approach is as follows. Let $Q(\omega, h)$ denote the true value of the h^{th} quantile of estimates at frequency ω. We seek the probability distribution of $P(Q(\omega, h)|data)$: the distribution of the true value of the h^{th} quantile, given the observed data. To find this, we use a Bayesian approach with the conservative choice of an uninformative (flat) prior for the power spectral value. Using Bayes' theorem, we can express $P(Q(\omega, h)|data)$ in terms of the probability of drawing the data from a distribution with known h^{th} quantile, or $P(data|Q(\omega, h))$:

$$P(Q(\omega,h)|data) = P\big(data\big|Q(\omega,h)\big) * P\big(Q(\omega,h)\big)/P(data)$$
$$\propto P(data|Q(\omega,h)) \qquad [6]$$

To be consistent with our prior reasoning for implementing a two-tiered approach, we implement the Bayesian approach to confidence intervals by considering the data to be the set of spectral estimates $\hat{S}_b(\omega)$ derived from each segment (taking the mean of the tapered estimates within each segment, equation [4]).

We then use order statistics to compute $P(data|Q(\omega, h))$. Specifically, we re-label each $\hat{S}_b(\omega)$ as $Y_1, \ldots, Y_B$, where Y_1 is the smallest ranked value in $\{\hat{S}_b(\omega)\}$. We also denote $Y_0 = -\infty$ and $Y_{B+1} = +\infty$, as this will allow us to account for the possibility of $Q(\omega, h)$ lying below the smallest ranked value or above the largest. The probability that $Q(\omega, h)$ lies between Y_i and Y_{i+1} is equal to the probability that exactly i of the $\hat{S}_b(\omega)$ estimates are below this quantile, and $B - i$ are above it. Since

the chance of any single estimate lying below the h^{th} quantile of the estimates is exactly h, this probability is determined by the binomial distribution:

$$P(Y_i < Q(\omega, h) < Y_{i+1}) = \binom{B}{i} h^i (1-h)^{B-i}. \quad [7]$$

Thus, to ensure that the probability that $Q(\omega, h)$ lies between two ordered values, Y_l and Y_m (where $m > l$) is at least $1 - \alpha$, we need to find indices l and m for which

$$\sum_{i=l}^{m-1} P(Y_i < Q(\omega, h) < Y_{i+1}) \geq 1 - \alpha \quad [8]$$

We choose the intervals in descending order of probability to determine the smallest number of intervals in eq. [8] for a given coverage $1 - \alpha$. The union of these intervals is the desired confidence interval. **Figure 2.1** illustrates this procedure for $\alpha = 0.05$ (i.e., 95% confidence intervals).

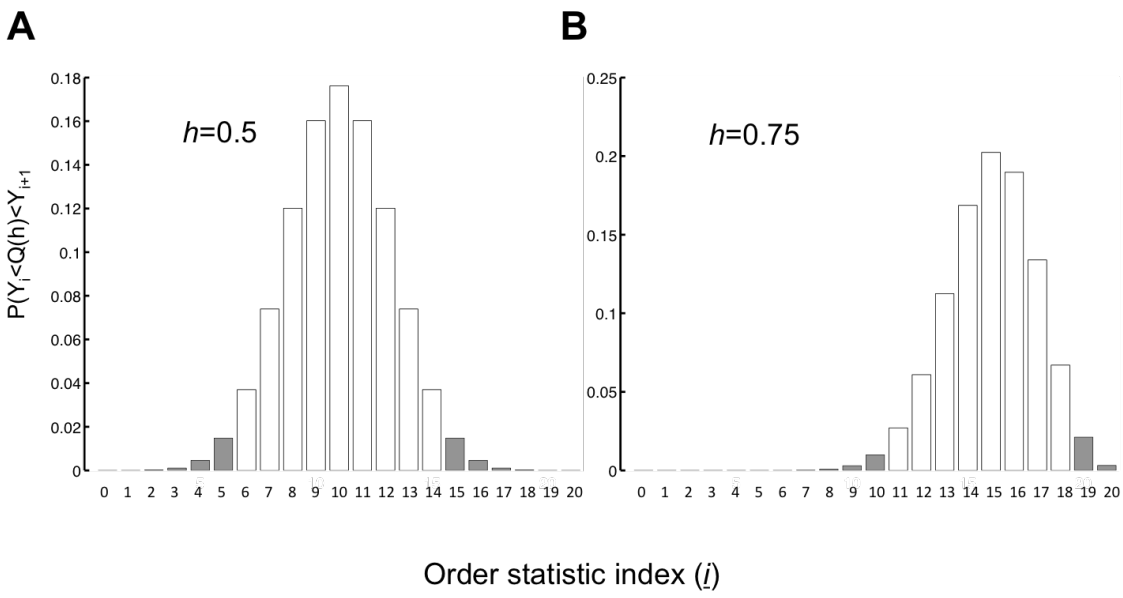


Figure 2.1: Procedure for determination of confidence intervals based on order statistics. The bar graphs show the distribution of probabilities of the true h^{th} quantile falling between the i^{th} and $(i+1)^{\text{th}}$ order statistic for a set of $B = 20$ values. The first and last bars ($i = 0$ and $i = B$) indicate the probability of the true quantile value falling in the intervals $(-\infty, Y_1]$ and $[Y_B, \infty,)$, respectively, where $Y_1, \ldots, Y_B$ are the order statistics corresponding to the B spectral estimates. The white bars indicate inter-order-statistic intervals whose probabilities sum up to $1-\alpha$, representing the region between the $1-\alpha$ confidence intervals.

Note that the confidence interval provided by the above procedure for a coverage $1-\alpha$ typically also applies to coverage factors somewhat larger than $1-\alpha$. This is because the upper and lower confidence intervals are tethered to discrete values (the observations Y_b), so the confidence bounds that satisfy equation [8] typically also satisfy it for smaller values of α as well. The relationship of the predicted coverage factor to the number of samples is shown in **Figure 2.2A** (for $\alpha = 0.05$ and $h = 0.5$). Note also that if the number of trials is sufficiently small, then it may be necessary to include the intervals $(-\infty, Y_1]$ and $[Y_B, \infty)$ in order to satisfy equation [8]. **Figure 2.2B** shows the minimum number of trials required to have finite confidence intervals, as a function of h.

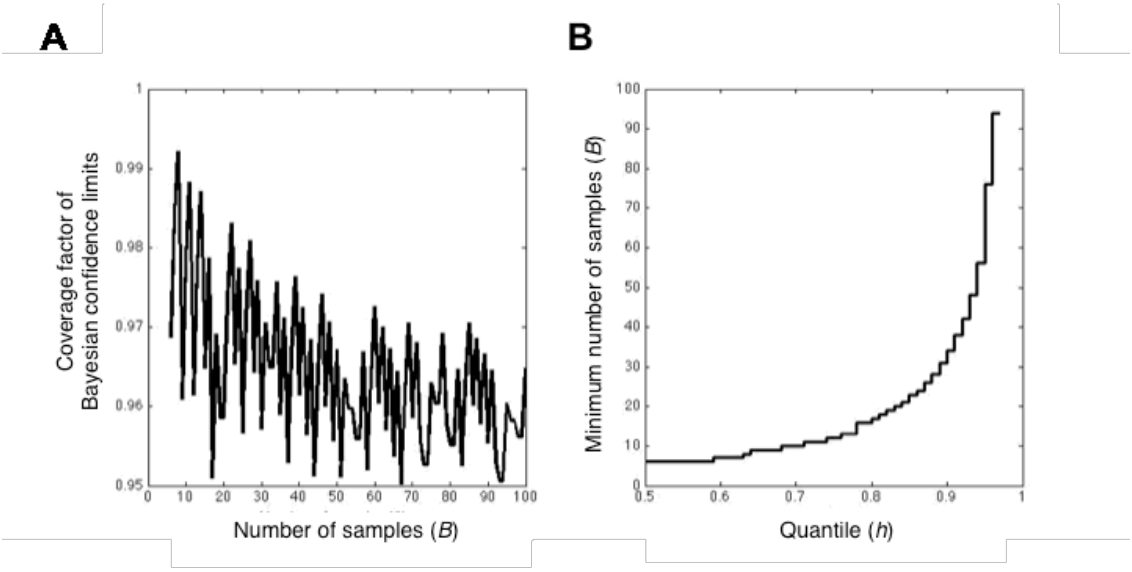


Figure 2.2: A. The nominal coverage factor of a Bayesian confidence interval exceeds $1-\alpha$, and this excess depends on the number of samples. Here this relationship is shown for the median ($h = 0.5$) estimator. B. The minimum number of samples necessary in order to obtain finite 95% confidence intervals for a range of quantile values.

For comparison purposes, we also calculated confidence intervals using two resampling methods: jackknife and bootstrap. The jackknife confidence interval is computed by pooling together all estimates from all tapered trials (for a total of BK estimates) and generating BK subsets of size $BK - 1$ by dropping one tapered trial from each. The standard or robust estimator of central tendency (mean for the standard method; quantile for robust) is applied to each subset. The standard deviation is calculated and significance is determined according to the t-distribution with $BK - 1$ degrees of freedom. (Note that for the standard multitaper method, this is the default Chronux implementation.) Bootstrap confidence intervals were calculated from 10,000 resamplings, with replacement, of the BK estimates, and again applying the appropriate estimator of central tendency (mean or quantile) to each. The $\alpha/2$ and $1 - \alpha/2$ quantiles of this distribution are used for the lower and upper confidence bounds. We emphasize that jackknife-based confidence intervals, which we have included in the provided Chronux extension, are only intended for purposes of comparison with the Bayesian confidence intervals (see below).

2.2. Method Validation

We applied the above procedures to (1) a synthetic signal of a known power spectral distribution corrupted by noise, and (2) an EEG record from a human test subject, with the typical artifacts of clinical recordings.

2.2.1. Simulated data

The simulated data consisted of a Gaussian signal of known power spectral distribution, to which we added a controlled amount of simulated artifact. The signal was synthesized from random-phase, Gaussian-distributed Fourier components, whose mean power was proportional to $1/\omega$ (over the range $1/3$ to 100 Hz), where ω is the frequency. Noise, which was added in the time domain, consisted of bursts of Gaussian signal with a flat power spectrum. The burst length was 0.5 seconds (while samples were simulated to be 3 seconds in duration), and bursts were inserted at Poisson-distributed intervals. We studied how the performance of the two methods varies with different values for the proportion of contaminated data segments in the data set. Sample data, both with and without artifact, are shown in **Figure 2.3**. Note that since the power spectrum of the underlying signal was $1/\omega$, we anticipate that at sufficiently high frequencies, the noise bursts will lead to unacceptable corruption of the spectral estimates.

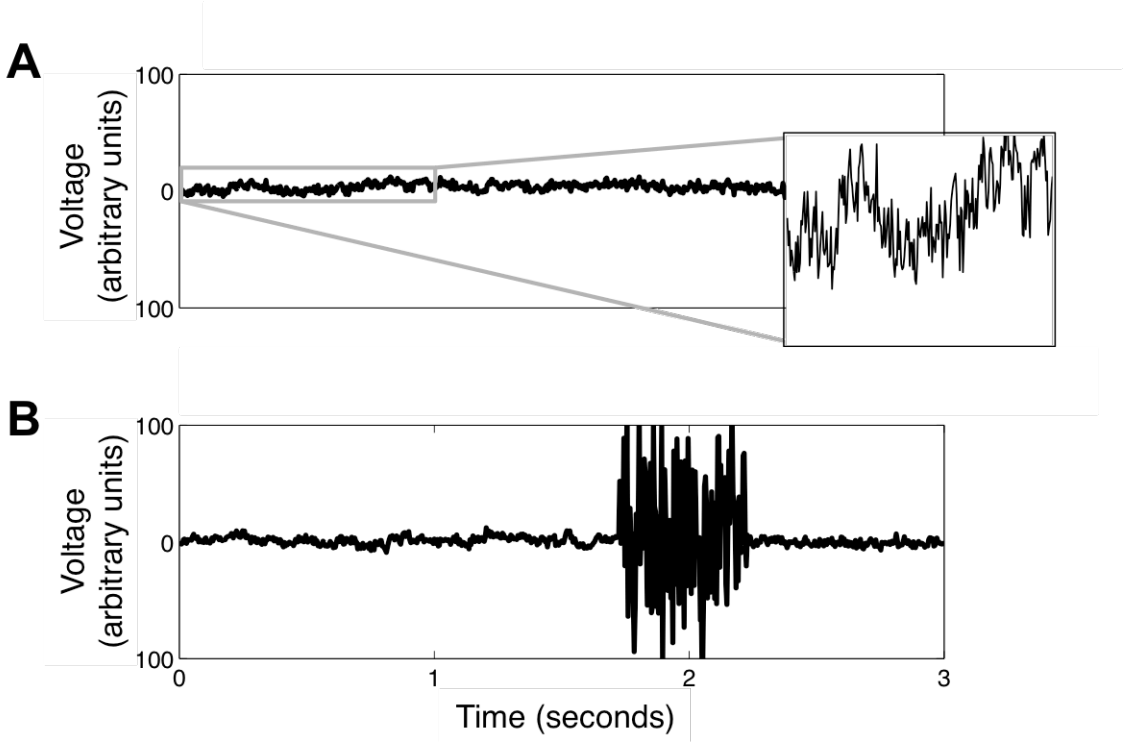


Figure 2.3: Representative simulated signal segments. A: three seconds of a simulated signal with a $1/\omega$ power spectrum. Inset enlarges a portion of the trace. B: three seconds of simulated artifact-containing signal, containing a white noise burst of 0.5-second duration.

2.2.2. Human data

The human EEG data were obtained from a healthy control subject (23 year old male). Data were recorded with an FS128 head box using an augmented 10/20 montage. The sampling frequency was 250 Hz. Input impedances were $\leq$5 KΩ. The subject was awake during testing, and generated spontaneous movement artifacts and eye blinks. For analysis, the signal was cut into three-second segments. Low-pass and high-pass filters of 0.01 and 100 Hz, respectively, and a 60 Hz notch filter, were applied. Segments were labeled as "artifact-containing" if they contained EMG, eye-blinks, or other artifacts, as determined by visual inspection of the EEG and simultaneously-recorded video (carried out by an experienced EEG analyst); and "clean" otherwise (**Figure 2.4**). To generate datasets with various levels of contamination by artifact, we drew segments randomly from these two subgroups.

Human subject participation was approved by the Institutional Review Board, and was consistent with the Declaration of Helsinki.

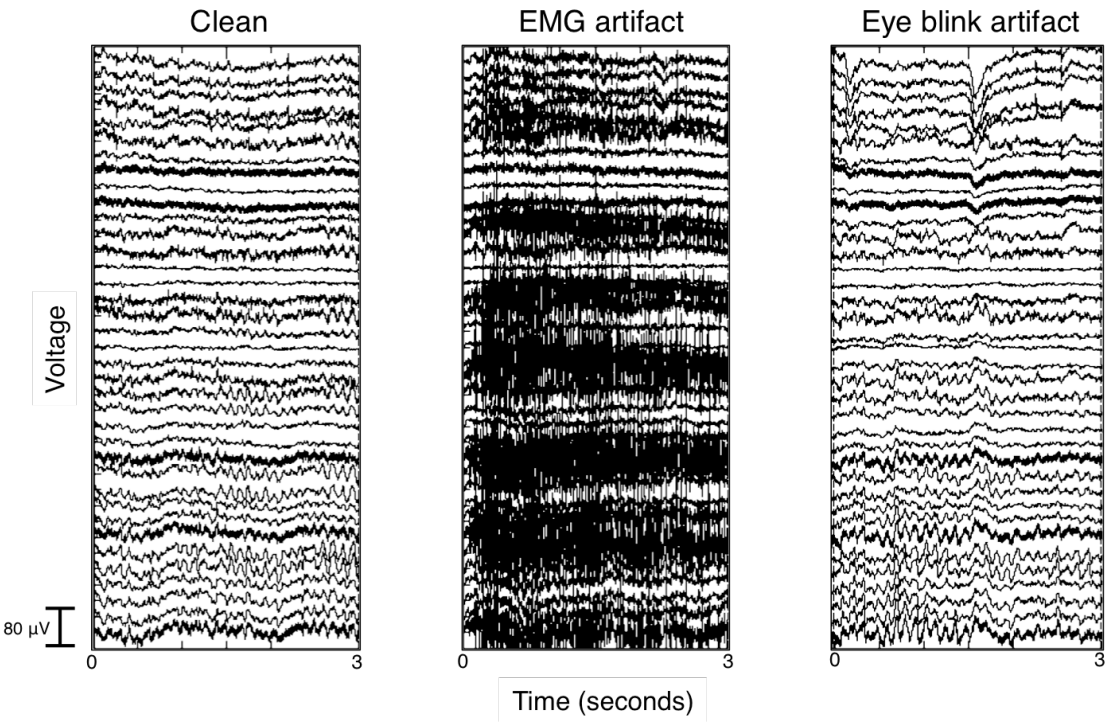


Figure 2.4: Typical segments of recorded human EEG in the time domain. The first panel shows an example of clean (artifact-free) signal, taken to consist predominantly of neural signal. The second and third panels show examples of EMG (electromyogram) and eye blink artifacts.

3. Results

We first compare the standard multitaper method and the robust method for simulated signals. Because the true underlying spectrum is known, this allows rigorous assessment of accuracy and the coverage factors of the estimated confidence intervals. We then apply the standard and robust methods to a sample of human EEG, and show that the robust method is less sensitive to typical EEG artifact encountered in clinical recordings.

3.1. Simulated EEG results

Figure 2.5 compares power spectral estimates via the standard and robust multitaper methods on a simulated EEG signal. Each method was applied to three data sets that differed in the average number of noise-contaminated segments per data set: a noise-free dataset, and datasets with an average of 25% and 50% noise-containing segments, respectively (see Methods for details). For the standard method, when artifacts were present the expected high-frequency decline of the power spectrum is corrupted by the flat spectrum of the noise bursts. This shows that the artifact significantly affects the estimated spectrum. In comparison, results from the robust method shown on the right panel reveal that even over frequencies at which the power spectrum of the standard estimate is dominated by the noise bursts, the robust estimate reflects the underlying signal's spectrum. In sum, in a data set where outliers significantly affect spectral estimates from the standard method, the estimate from the robust method can capture the underlying spectrum.

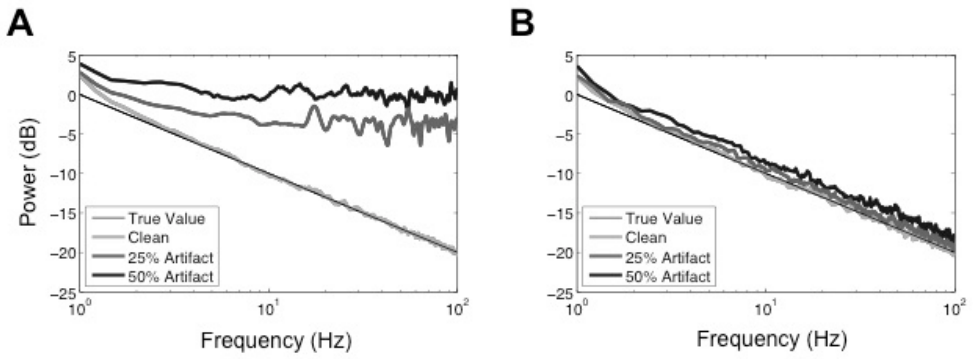


Figure 2.5: Power spectral estimates (left: standard method, right: robust method) for simulated datasets with 0%, 25%, and 50% artifact-containing segments, respectively. The thin sloping black line is the true value of the power spectrum ($1/\omega$). Spectra were estimated with 5 tapers and a time-bandwidth product of 3.

The simulated data allowed for an assessment of confidence interval estimation methods (**Figure 2.6**)**.** The first column of the Figure shows results for the standard approach, i.e., the standard multitaper estimates with jackknife error bars. As expected, when no noise is present (top panel), the spectral estimates are close to the true value, and confidence interval coverage is approximately 95%. When noise is added, the spectrum is upwardly biased by the noise, and the coverage factors drop. The last column of the figure shows that for the robust method, the error bar coverage determined by the Bayesian procedure described in Methods remains at approximately 95%, even when noise is added.

The rest of the columns in **Figure 2.6** show the characteristics of confidence intervals that are provided by alternative approaches. Specifically, the second column shows that when a jackknife procedure is applied to the robust method, the confidence intervals have a lower-than-veridical coverage factor, and they are also more irregular than the Bayesian confidence intervals. With an implementation of the bootstrap (third and fourth columns), neither confidence intervals on the standard method nor on the robust method achieve the coverage of the Bayesian confidence intervals. As mentioned in Methods, the failure of confidence intervals based on resampling is not surprising, since the median depends in a non-smooth fashion on the data. However, even under these circumstances – and when artifacts are not explicitly removed -- the Bayesian confidence intervals are approximately veridical.

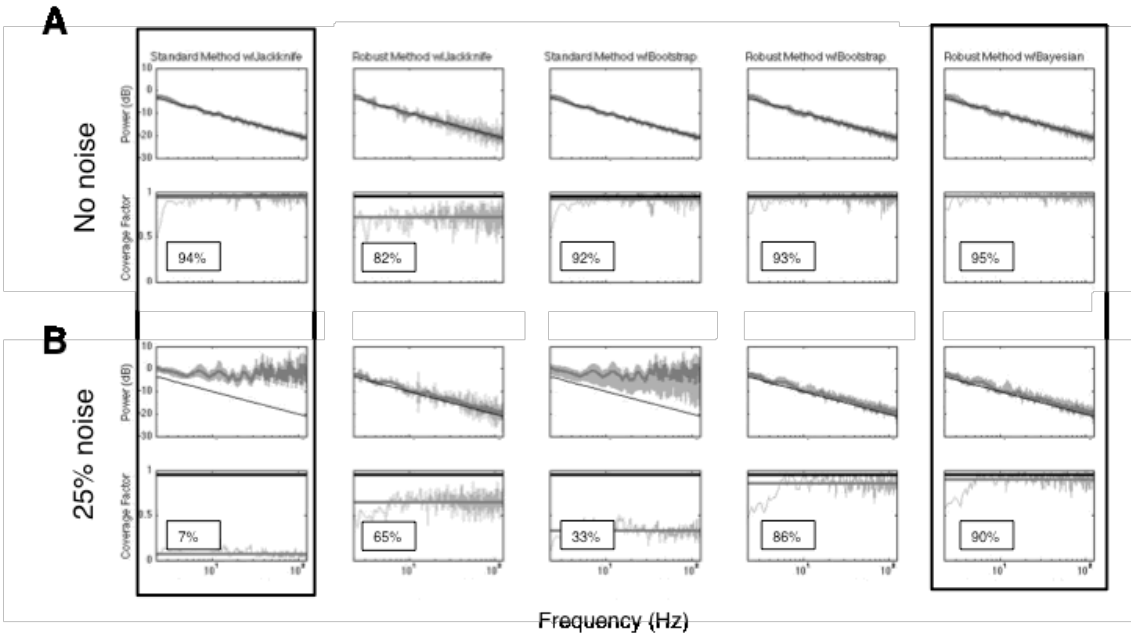


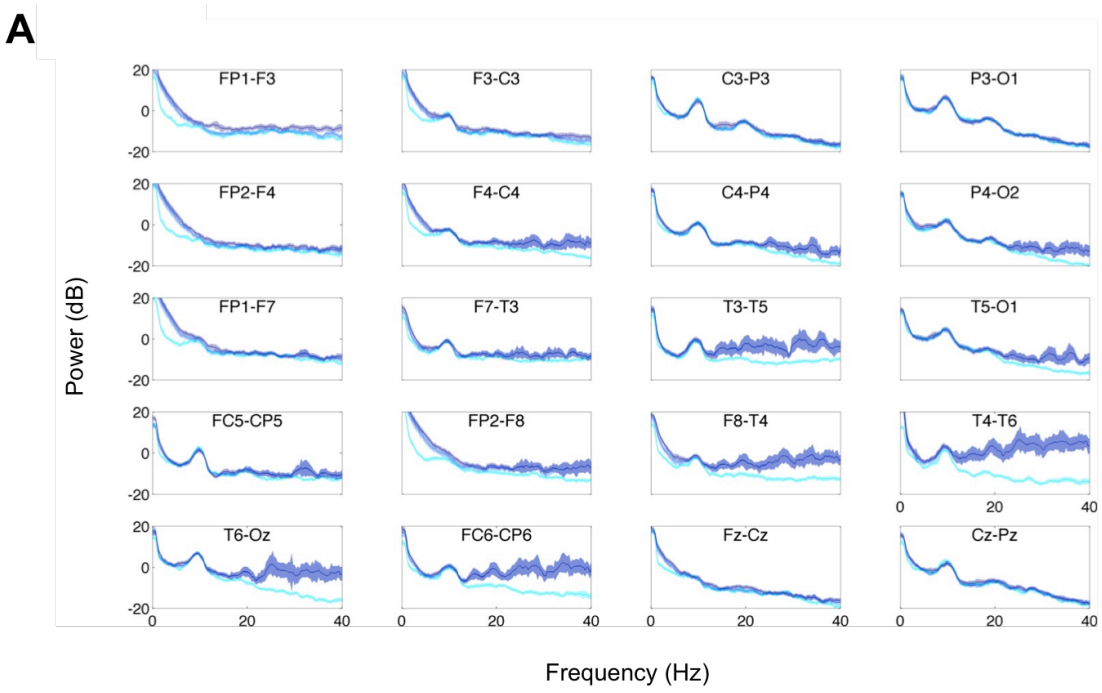
Figure 2.6: Coverage factors of confidence intervals for simulated data obtained via several methods, for simulated data. Simulations as in Figure 3: spectra are estimated for data sets of 20 samples with 5 tapers and a time-bandwidth product of 3. A: artifact-free data; B: 25% of segments on average contaminated by artifact. Left column is the standard multitaper estimate of power spectrum, with jackknife error bars. Right column is the robust multitaper estimate of power spectrum with Bayesian error bars. Three middle columns use the jackknife with the robust method, or bootstrap with either the standard or robust method (see Methods for details). In the plots of coverage factor vs. frequency, the thin gray line represents the coverage factor at each frequency, averaged over 30 simulations; the black line indicates 95% coverage, and the thick gray line represents the average coverage for each method (also indicated by the number displayed on each graph). Note that the robust method with Bayesian error bars is the only one that achieves close to 95% coverage when artifact contamination is present.

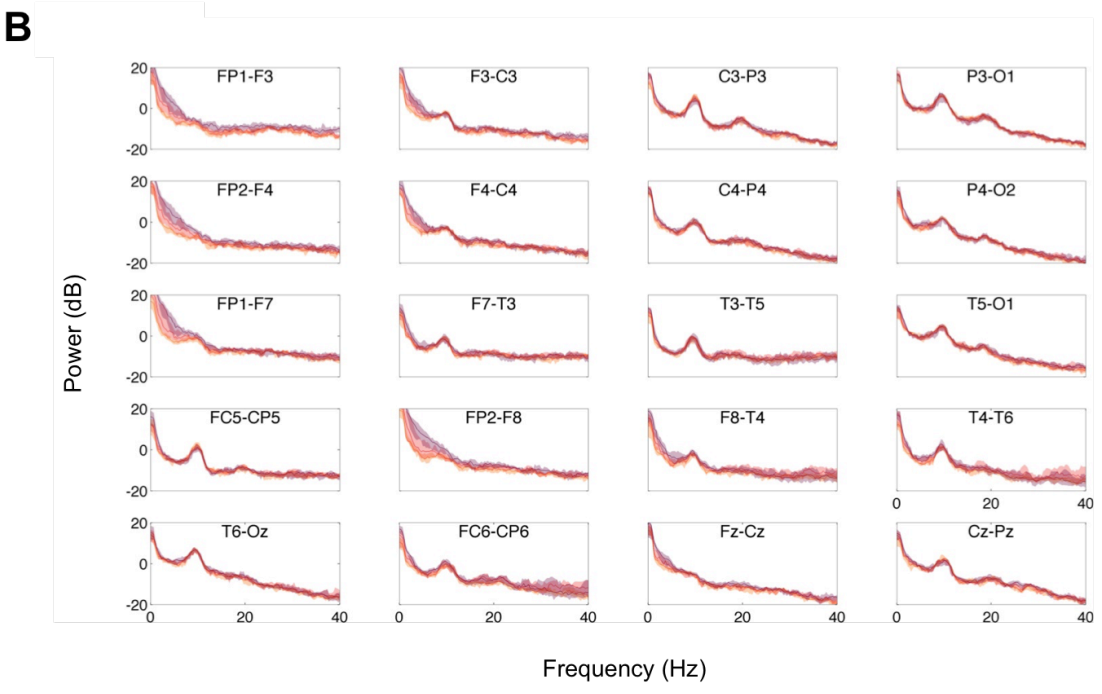
3.2. Human EEG results

Figure 2.7 shows that the robust method can be successfully applied to human EEG data. **Figure 2.7A** compares spectra determined by standard analysis of hand-cleaned EEG segments with a set of EEG segments for which 25% or 50% contained visible artifact. As expected, there were substantial deviations of the spectra obtained without hand cleaning. These included deviations at low frequencies in frontal channels, presumably due to eye movement artifact, and deviations and at high frequencies in frontal and posterior channels, presumably due to EMG artifact. **Figure 2.7B** shows that these deviations, especially at high frequencies, are largely eliminated when the robust method is used.

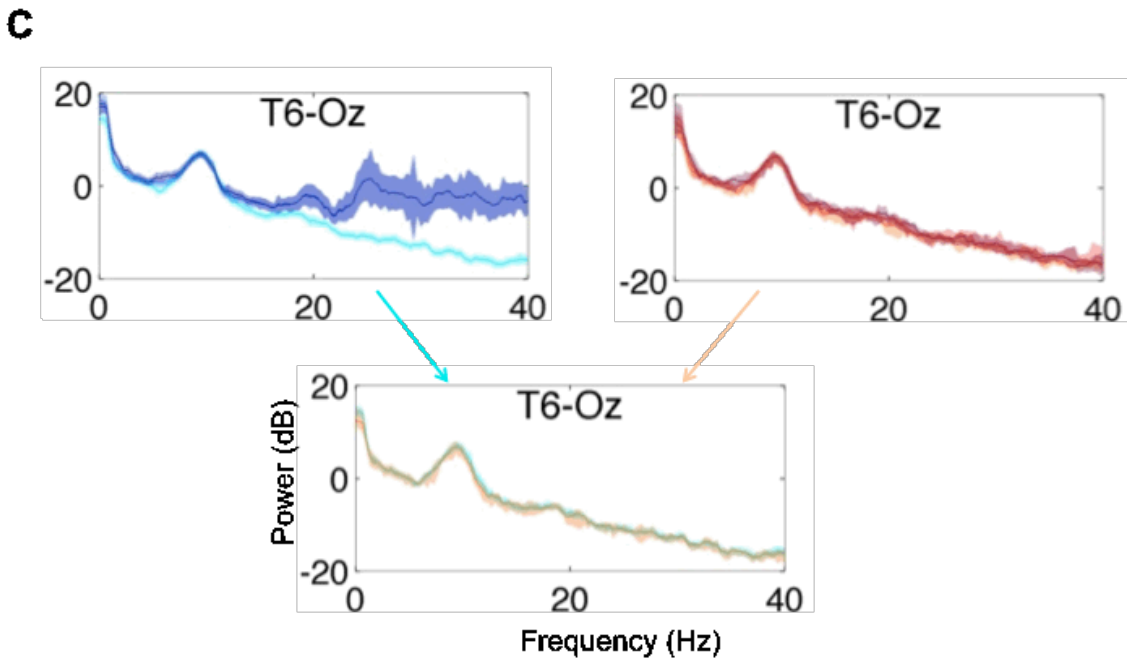
Figure 2.7C shows that on cleaned data the robust method gives results that are very close to that of the standard method. Thus, the resistance of the robust method to corruption by artifact is not at the expense of a distortion of the result.

Figure 2.7: Power spectral estimates (A: standard method, B: robust method) for 20 channels of EEG from a human subject. Spectra were calculated using 5 tapers and a time-bandwidth product of 3. Light color indicates 0% artifact-containing segments; intermediate color indicates 25% artifact-containing segments; and dark color indicates 50% artifact-containing segments. Color bands indicate 95% confidence intervals. C. Detailed comparison at channel T6-Oz. Note that for artifact-free data (lower panel), the two methods give similar results.


A
FP1-F3
F3-C3
C3-P3
P3-O1
FP2-F4
F4-C4
C4-P4
P4-O2
FP1-F7
F7-T3
T3-T5
T5-O1
FC5-CP5
FP2-F8
F8-T4
T4-T6
T6-Oz
FC6-CP6
Fz-Cz
Cz-Pz
Power (dB)
Frequency (Hz)


B
FP1-F3
F3-C3
C3-P3
P3-O1
FP2-F4
F4-C4
C4-P4
P4-O2
FP1-F7
F7-T3
T3-T5
T5-O1
FC5-CP5
FP2-F8
F8-T4
T4-T6
T6-Oz
FC6-CP6
Fz-Cz
Cz-Pz
Power (dB)
Frequency (Hz)

Continued figure 2.7:



4. Discussion

Above we have shown that a simple modification of the standard approach to spectral estimation – substituting a robust estimator for the mean across segments – can substantially improve spectral estimates of EEG signals in the presence of artifact. The basic rationale is that robust estimators are insensitive to outliers, and many sources of artifact behave as outliers. When combined with the multitaper method (Thomson, 1982), key advantages of the latter are retained: spectral leakage is minimized, and reliable confidence intervals can be estimated.

The proposed data-driven approach reduces the reliance on removal of artifact by other means. This has several advantages: with less preprocessing, fewer data will be discarded, potentially enabling the capture of subtle EEG dynamics, as well as the retention of statistical power. A reduced reliance on preprocessing methods also has the benefit of reducing the dependence on *ad hoc* or subjective methods of artifact identification, and may also accelerate the data-processing pipeline.

It is worth noting that the robust methods described here are computationally efficient. Since the MATLAB implementations of `median()` and `quantile()` (used for the power spectrum) and `sort()` (used for Bayesian confidence intervals) have approximately linear runtime even for 10^8 segments, the robust method retains the linear asymptotic runtime of the standard multitaper approach.

4.1. Caveats

Because the robust method works by separating signals that are pervasive (e.g., those reflecting background state) from those that are infrequent, it may discard an intermittent signal of neural origin – such as paroxysmal activity – as artifact. The robust method works specifically because it removes outliers. Outliers include many sources of artifact but can also include neural-origin EEG activity that is infrequent.

Although this method greatly improves spectral estimates for certain data sets, it should not be treated as a panacea for all analysis-limiting noise. Since quantile estimators have a breakdown point of ≤50%, this method may not show any improvement over the standard analysis pipeline for constant or frequent noise that affects the same frequency range in most or all segments, such as 60 Hz line noise from electronics in the environment or frequent muscle tics. In these cases, alternatives such as notch filtering or artifact removal by hand must be used.

For clarity we tested the method here in the absence of other artifact removal techniques. However, there are benefits to other techniques for removing outliers, such as ICA or hand cleaning guided by video assessment of the subject's movements. Combining artifact removal techniques with the robust method may be more effective than either approach on its own.

4.2. Extension to multichannel analysis

The utility of robust estimators as applied to the EEG spectrum suggests that robust methods will also be useful in the multichannel domain. In this context, robust estimators of shape, such as the minimum volume ellipse (MVE) (Rousseeuw and

Leroy, 1987) or the minimum-covariance determinant (MCD) (Hubert and Debruyne, 2010), could be used to estimate cross-spectra, much as the quantile-based estimators here characterize power. While this specific robust approach appears to be as yet unexplored, a previous study has shown that using the median instead of the mean improves multi-taper coherence estimates (Wong et al., 2011). A full-fledged robust estimator of shape could provide phase information as well.

Acknowledgments

We thank Mary Conte for collecting and providing the EEG data, Tanya Nauvel for hand-cleaning the data, and Jonathan Drover for discussions on the theoretical background in this manuscript. We also thank Tanya Nauvel for her helpful comments on the manuscript and Andrew Golfine for code testing.

Support was provided by the Tri-Institutional Training Program in Computational Biology and Medicine (via NIH training grant 1T32GM083937 to TM) and the James S. McDonnell Foundation (Nicholas D. Schiff, PI).

Appendix: Scale factor

As mentioned in the main text, the raw quantile of a set of spectral estimates is expected to be proportional to the power, but not equal to it. The proportionality constant is dependent on the quantile h, the number of degrees of freedom d of the underlying χ^2 distribution describing the expected distribution of spectral estimates for Gaussian data, and the number of samples (segments), B. Here we derive this scale factor $C(h, d, B)$ by determining the h^{th} quantile of the tapered spectral estimates, $\hat{S}_b(\omega)$, for a Gaussian signal of unit power.

For most frequencies ω, with the exception of $\omega = 0$ or the Nyquist frequency, Fourier estimates are complex numbers. When ω is greater than the bandwidth of the tapers, the real and imaginary components are approximately independent and of equal variance (Percival and Walden, 1993), so for $\{\hat{S}_b(\omega)\}$ the power is distributed as the sum of $2K$ squares of Gaussian-distributed quantities, where K is the number of tapers. For $\omega = 0$ or the Nyquist frequency the Fourier estimates are real, so the power of the $\{\hat{S}_b(\omega)\}$ is distributed as sum of K such quantities. Therefore $\{\hat{S}_b(\omega)\}$ at a particular frequency ω is distributed as χ^2/d where χ^2 has d degrees of freedom, and $d = K$ or $d = 2K$. The proportionality between the quantiles of this distribution and the mean is needed to convert the quantile estimate into an estimate of power.

To compute the expected value of a quantile, we use the strategy shown in **Figure 2.8**. We first find a monotonic transformation from the uniform distribution

on the interval [0,1] into the chi-squared distribution of spectral estimates. Because the transformation is uniform, the rank-order of the samples drawn from the chi-squared distribution corresponds to the rank-order of the corresponding values in the uniform [0,1] distribution. Therefore, we can take the expected distribution of the h^{th} quantile in the uniform distribution, and transform it back into the chi-squared distribution to determine $C(h, d, B)$.

To determine this transformation, we note that for an arbitrary distribution with probability density function (pdf) $q(x)$, the cumulative distribution function (cdf) $s(x)$ is given by

$$s(x) = \int_{-\infty}^{x} q(z)dz.$$

This can be rewritten as

$$\frac{ds(x)}{dx} = q(x).$$

The cumulative distribution function, by definition, is uniformly distributed between 0 and 1. With $q(x) = \chi_d^2(xd)$, then $x = \frac{1}{d} chi2inv(s, d)$ is the desired transformation between a uniformly-distributed quantity, s, and the spectral estimates, where $chi2inv(s, d)$ is the inverse cumulative chi-squared probability density function at s with d degrees of freedom.

We now apply order statistics to the uniformly distributed variable s, which ranges between $[0,1]$. The distribution of the $(k+1)^{\text{th}}$-ranked value for $N + 1 = B$ draws from the uniform distribution is given by

$$p_{k,N}(s) = \frac{1}{beta(k+1, N-k+1)} s^k (1-s)^{N-k},$$

where $beta(u,v)$ is the beta-function, $beta(u,v) = \int_0^1 t^{u-1}(1-t)^{v-1}dt$ (David and Nagaraja, 2003).

Transforming back to the distribution of spectral estimates, we find:

$$q_{k,N}(x)dx = p_{k,N}(s)ds,$$

where

$$q_{k,N}(x) = p_{k,N}(s)\frac{ds}{dx} = \frac{1}{beta(k+1, N-k+1)} s^k (1-s)^{N-k}\frac{ds}{dx}.$$

The expected value of this quantity is therefore:

$$\langle x \rangle = \int_0^\infty x(s) q_{k,N}(x)dx = \frac{1}{beta(k+1, N-k+1)} \int_0^\infty x s^k (1-s)^{N-k} \left(\frac{ds}{dx}\right) dx$$

$$= \frac{1}{beta(k+1, N-k+1)} \int_0^1 \frac{1}{d} chi2inv(s,d) s^k (1-s)^{N-k} ds$$

When the quantile h falls exactly on a sample, i.e., when $h = \frac{k+1}{N+1} = \frac{k+1}{B}$, this is $C(h,d,B)$. When the quantile h falls between two samples, $C(h,d,B)$ is determined by interpolating this value between two adjacent values of k. By default, the code uses the MATLAB convention for quantile interpolation, i.e., a weighted average of the values at the two adjacent values of k.

As $B = N + 1$ increases, the above result takes a simple asymptotic form, since the integrand factor $s^k(1-s)^{N-k}$becomes concentrated at $s = \frac{k+1}{N+1} = h$. In this limit,

$$\lim_{B\to\infty} C(h,d,B) = \frac{1}{d} * chi2inv(h,d).$$

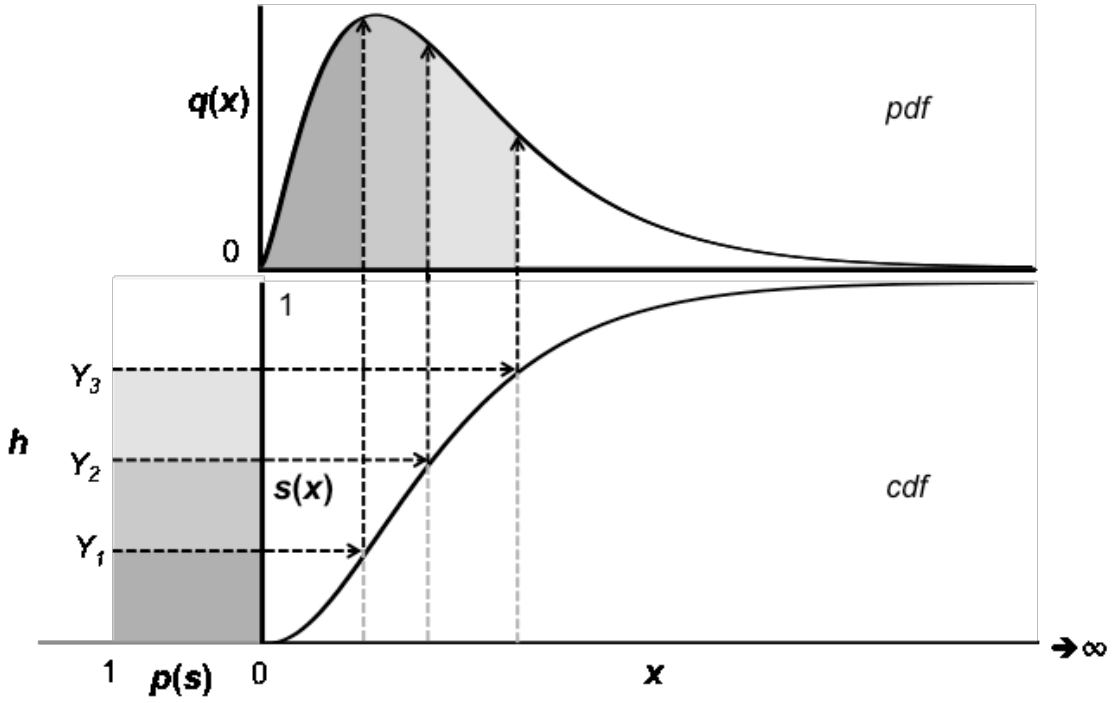


Figure 2.8: Determining the order statistics of samples drawn from a non-uniform distribution whose PDF is given by $q(x)$. First, $s(x)$, the CDF associated with $q(x)$, is used to map samples drawn from a uniform distribution on the $[0, 1]$ interval into samples drawn from $q(x)$. The function $s(x)$ is necessarily monotonic, so this mapping does not change the rank order of samples. Order statistics for the uniform distribution, Y_i, are then mapped via $s(x)$ into the domain for $q(x)$.

CHAPTER 3: ROBUST COMPARISON OF EEG POWER SPECTRA

Building on the results from chapter 2, we expanded the robust multitaper power estimation toolbox to include an approach for comparing two median power spectra. We developed a novel method based on the Bayesian approach to confidence intervals on the median shown in chapter 2, and compared it to established significance tests. The body of this chapter is a pre-submission version of a manuscript.

Introduction

The electroencephalogram (EEG) is a noninvasive measure of brain activity in widespread clinical and research use. It records voltage fluctuations from the surface of the scalp, which includes signals from cortical circuitry. However, a drawback of EEG is that, along with this recording of brain activity, the recorded voltage fluctuations also contain electrical signals from non-cortical sources, including (but not limited to) muscle movement, eye blinks, electrode movement, and electrical noise from the environment. These artifactual signals typically appear as high-amplitude voltage fluctuations relative to the neural signals, overpowering and obscuring them, and posing a barrier to quantitative analysis. Here, we focus on a

statistical strategy to mitigate the effects of these artifacts, in the specific context of comparison of EEG spectra.

The power spectrum has an important place in the characterization of the EEG because it directly corresponds to the correlation structure of neural activity, and therefore is informative about the dynamics of cortical populations. It is therefore often of interest to compare spectra from two different states: for example, during a task versus baseline, during different times of the day, before and after treatment for brain injury, etc. This can be done via a comparison test, such as the two-group test (Arvesen, 1969), which identifies the frequencies at which two spectra differ by a greater amount than would be expected by chance.

Outliers can produce misleading results in two ways. False positives can occur if the spectra being compared contain different levels or types of artifact: the comparison may show spectral differences merely because of differences in artifact content, even if there are no differences in the underlying spectra. False negatives can also occur, if true signal differences are obscured by large amounts of similar artifact. Moreover, a comparison may correctly indicate a difference in power, but be incorrect about which underlying spectrum is larger: this can happen if the smaller underlying spectrum is associated with larger-amplitude artifacts, Because outliers can lead to false positives, false negatives, and errors in the sign of the comparison, the difficulties associated with them cannot be solved merely by adjusting the stringency of a standard comparison, such as the two-group test.

In the case of artifacts that appear intermittently, cleaning the signal can be done by removing time segments that contain significant artifact (either by hand or using automated methods), or by removing spatial components containing the majority of the artifact via independent component analysis (ICA). Both methods are commonly used, and can be followed up with a standard multitaper calculation of the power spectrum (Thomson, 1982). A third option involves handling outliers at the level of spectral estimation, by using a robust estimator of location such as the median or quantile statistic instead of the standard multitaper approach (Melman and Victor, 2016). In practice this separates high-amplitude intermittent (non-neural) signals from the low-amplitude continuous (neural) signal, and estimates te spectrum of the latter, by taking the median over several windowed estimates. This third approach can be combined with pre-processing to remove artifact, either by hand or automatically.

In (Melman and Victor, 2016), the third approach was developed as both a stand-alone alternative to current standard techniques as well as one that can be combined with preprocessing for temporal or spatial removal of artifact. However, because such spectra are based on the median rather than the mean, comparing them by the parametric two-group test is no longer rigorously correct. This motivates the use of non-parametric tests—such as the rank-sum or permutation tests—to compare these robustly-estimated spectra. Here, we investigate the efficacy of this approach by comparing the parametric t-test and jackknife two-group test (Arvesen, 1969; Bokil et al., 2007) against the nonparametric approaches of permutation test of medians and

the Wilcoxon rank-sum test. We also develop a new Bayesian method for this purpose.

Methods

Typical EEG power spectral analysis involves segmenting the recording, denoted as $x(t)$, into m segments $\{x_1, x_2, \ldots, x_b, \ldots, x_B\}$ of length T, and applying a frequency decomposition. In this case, we apply the multitaper method [Thompson 1982], which (1) applies a series of k orthogonal windowing functions called Slepian tapers to each segment, (2) calculates the norm-squared of the Fourier components of the windowed segments, and (3) averages over the tapers:

$$S_b(\omega) = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{T}\left|\int_0^T e^{i\omega\pi}x_b(t)a_k(t)dt\right|^2\right)$$

where $S_b(\omega)$ is the power spectral estimate from data segment $x_b(t)$, as a function of frequency omega and $a_k(t)$ is the k^{th} taper. The total power spectrum is typically calculated as:

$$S(\omega) = \frac{1}{B}\sum_{b=1}^{B} S_b(\omega)$$

The Melman Victor 2016 robust approach uses the median instead:

$$S(\omega) = median(S_b(\omega))$$

Two compare two power spectra, we compare the individual $S_b(\omega)$spectral estimates from one dataset with the $S_d(\omega)$ values from the second dataset. We want to test whether the centers of both distributions--either the mean or the median--are equal.

Two-distribution comparison tests

We applied four previously established comparison tests to the test data. As parametric methods for comparison, we used the student's t-test and Arvesen's (1969) jackknife two-group test (as adapted for EEG power spectra by Bokil et al. (Bokil et al., 2007)). We compared these against the non-parametric rank-sum test and the permutation test on medians, and also against a new test, the Bayes box test, described below.

The Bayes Box Test

The Bayesian approach to testing the difference of medians is a third non-parametric approach; we develop it below and discuss it more fully in "Discussion". The goal is to construct the posterior distribution of the difference of medians. We restate the problem of comparing two data sets' quantiles as follows: given two

datasets, X of size m and Y of size n, how different are the observed medians? We sort each data set and assign order statistics so that the i^{th} ranked value of X is labeled x_i and the j^{th} ranked value of Y is labeled y_j. Consistent with the non-parametric nature of the test, we assume that each dataset's prior distribution on its respective median is flat in the space of the order statistics, and therefore bounded. Note that, as is the case for the spectral estimation method developed in Chapter 2, the prior is defined in the space of the order statistics, not in the space of the data.

The posterior distribution of the median of one dataset is given by the binomial distribution: as mentioned in (Melman and Victor, 2016), each x_i lies below the median (the 0.5^{th} quantile) with a probability $h = 0.5$, and above it with a probability $1 - h = 0.5$. Thus each sample corresponds to an independent trial with probability 0.5. Therefore the probability of the median lying between the i^{th} and $(i+1)^{\text{th}}$ ordered values is given by the binomial distribution:

$$P(x_i \leq median(X) \leq x_{i+1}) = \binom{m}{i} * 0.5^i * (1 - 0.5)^{m-i} = \binom{m}{i} * 0.5^m.$$

Assuming the two datasets are independent, the joint probability that $x_i < median(X) \leq x_{i+1}$ and $y_j \leq median(Y) \leq y_{j+1}$ is given by

$$P(b_{i,j}) = P(x_i \leq median(X) \leq x_{i+1}) * P(y_j \leq median(Y) \leq y_{j+1})$$

$$= \binom{m}{i} * \binom{n}{j} * 0.5^{m+n}$$

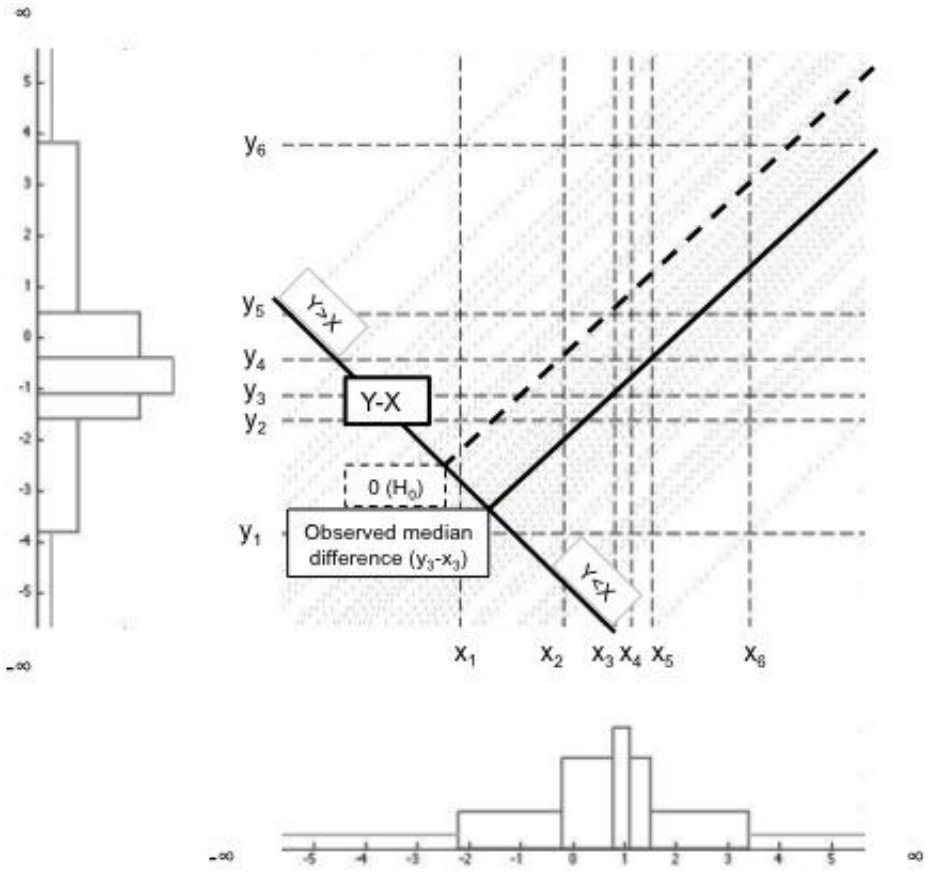
where $b_{i,j}$ denotes the box bounded by x_i and x_{i+1} on the x axis, and y_j and y_{j+1} on the y axis, containing the probability $P(b_{i,j})$.

We then calculate the joint distribution of the difference in medians as follows: calculate the m*n pairwise differences $y_j - x_i$ of each point y_j in Y with each point x_i in X. The probability density at $q_{0.5}(Y) - q_{0.5}(X) = d$, where d is some real value, is calculated by summing up all of the probability densities of regions in the joint space where $y_j - x_{i+1} \leq d \leq y_{j+1} - x_i$—i.e. the "boxes" in the joint probability plane through which the line $y = d + x$ passes (**figure 3.1**)—scaled by the proportion of the box probability, modeled as a uniform distribution within each interval. More explicitly: the projection of $b_{i,j}$ onto the line containing the ranked pairwise differences, denoted as Z, is a region bounded by $z_k = y_j - x_{i+1}$ and $z_l = y_{j+1} - x_i$. Note that the values z_k and z_l are not necessarily consecutive values in Z. Therefore, to determine the pdf in the space of the ranked pairwise differences, Z, we sum contributions from all of the boxes b_{ij} that contribute to $[z_q z_{q+1}]$, i.e. all i and j for which $y_j - x_{i+1} \leq z_q \leq d \leq z_{q+1} \leq y_{j+1} - x_i$, where the contribution from box $b_{i,j}$ at d is:

$$pdf_{b_{i,j}}(d) = P\left(b_{i,j}\right) * \frac{z_{q+1} - z_q}{\left(y_{j+1} - x_i\right) - \left(y_j - x_{i+1}\right)}.$$

The sum of these scaled contributions from each box through which d passes produces the pdf at d, and the distribution of all such intervals for all values of d constitutes a posterior distribution of the difference in medians (**figure 3.2**).

Figure 3.1. The joint distribution of the difference in medians of two datasets. Dataset X consists of ordered values $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ (along the x axis). Dataset Y consists of ordered values $\{y_1, y_2, y_3, y_4, y_5, y_6\}$ (along the y axis). Order statistics are spaced according to their corresponding values' spacing on the real number line. The posterior probabilities of each median respectively are shown on the axis; each is given by the binomial distribution with $n=6$ and $p=0.5$. The line Y-X represents the axis of pairwise differences of each point in Y with each point in X; larger pairwise differences y_j-x_i are farther toward the upper left direction of this line, and smaller y_j-x_i lie farther toward the lower right direction. The point 0 on this line corresponds to the line perpendicular to Y-X that passes through $y_j=x_i$—the null hypothesis (H_0), shown as a dashed line. The point median(Y)-median(X) corresponds to the line perpendicular to Y-X passing through the observed median difference. Each "box" in this plane defined by x_i, x_{i+1}, y_j, and y_{j+1} contains a probability corresponding to the product of X's and Y's posterior probabilities on their respective medians: $P(x_i<\text{median}(X)<x_{i+1})*P(y_j<\text{median}(Y)<y_{j+1})$. See figure 2 for constructing the posterior on the difference in medians from this figure.


∞
5
4
3
2
1
0
-1
-2
-3
-4
-5
-∞
y6
y5
y4
y3
y2
y1
Y-X
0 (H0)
Observed median difference (y3-x3)
x1
x2
x3
x4
x5
x6
-∞
-5
-4
-3
-2
-1
0
1
2
3
4
5
∞

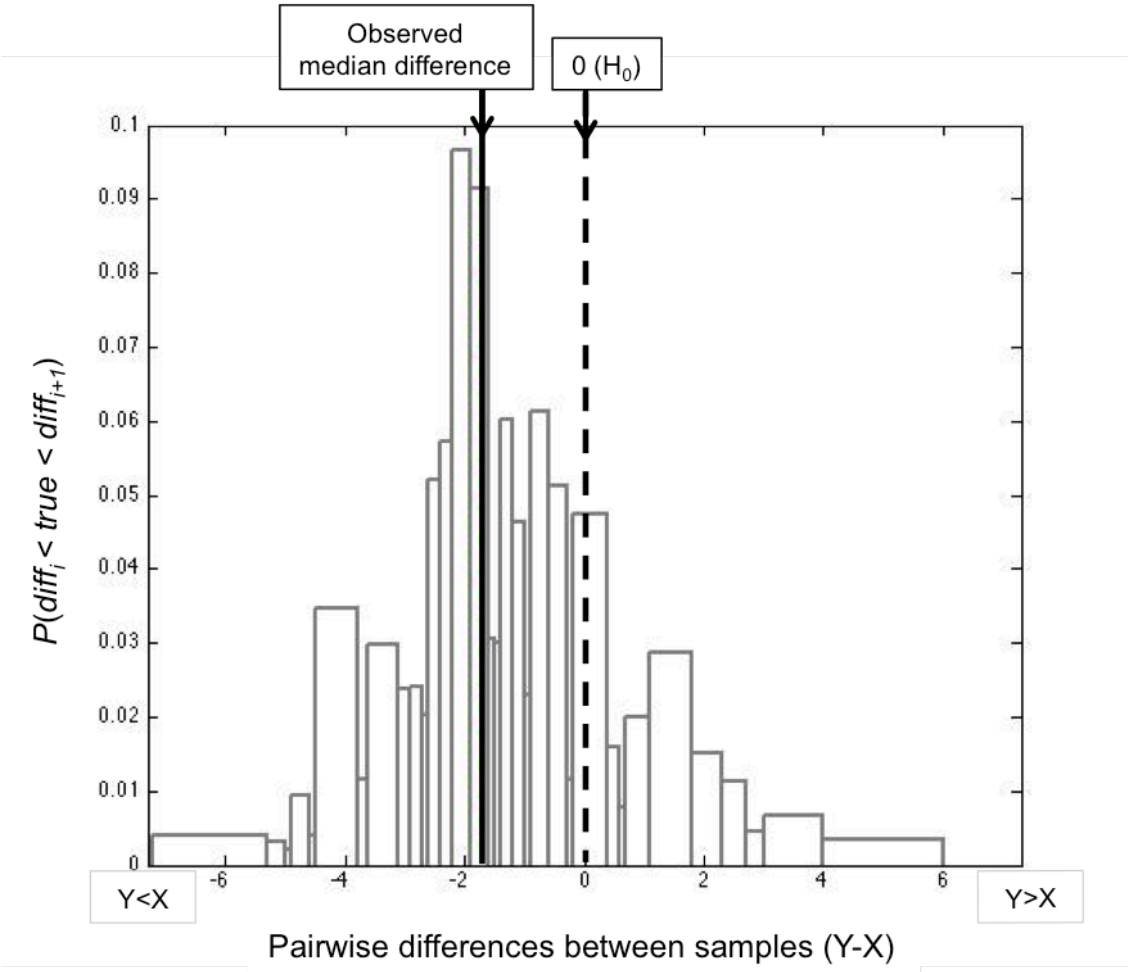
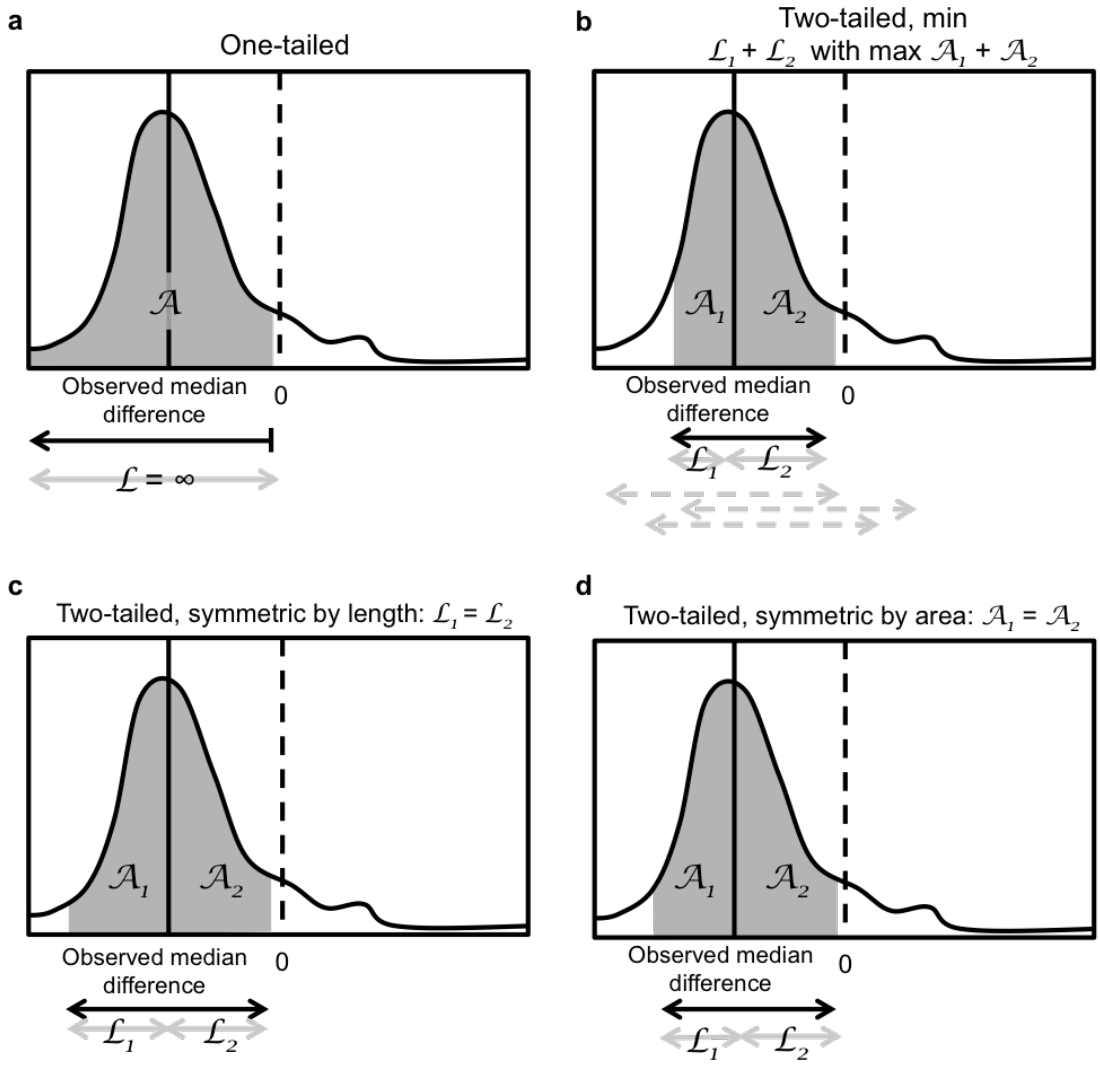


Figure 3.2. The posterior probability of the difference in medians. The x-axis is the line Y-X from figure 3.1. The set of pairwise differences between X and Y is the set DIFF, with ordered values {$diff_1$, $diff_2$, …, $diff_i$, …, $diff_{36}$} (with the example of |X|=6 and |Y|=6, there are 36 pairwise differences between data points). The probability contained in each interval d_i and d_{i+1} is given by the sum of the probability densities shown in figure 1 that lie between the lines y-x=$diff_i$ and y-x=$diff_{i+1}$.This is the posterior probability distribution of the difference in medians. The observed difference in medians (solid line) is shown relative to 0 (H_0) (dashed line).

With this posterior, we then obtain an interval of interest as follows: for a one-tailed test, the interval of interest is the curve to one side (either the left or the right, depending on the desired tail) of the observed median difference $q_{0.5}(Y) - q_{0.5}(X)$, excluding intervals that contain the observed value (**figure 3.3a**). For a two-tailed test, the interval of interest is the optimal interval ("optimal" is explained below) that contains the observed difference in medians but which excludes 0. The area under the curve outside of the interval of interest is the p-value of the test. The criterion of the "optimal interval" can be implemented in several ways: (1) largest most-compact interval: the interval of greatest compactness (area÷length) with the largest area that includes the observed median difference but excludes 0 (**figure 3.3b**); (2) largest symmetrical interval, by distance, on either side of the observed median difference that does not contain 0 (**figure 3.3c**); or (3) largest symmetrical interval, by area, on either side of the observed median difference that doesn't contain 0 (**figure 3.3d**). For symmetry (costs #2 and #3), implementation is straightforward since the endpoints are fixed. For optimal total interval length (strategy #1), implementation is carried out as follows: (1) list intervals by area and length, listed as $[(A_1, L_1), (A_2, L_2), \ldots (A_k, L_k), \ldots (A_{m*n}, L_{m*n})]$ so that $A_k$denotes the k^{th} ordered area and $L_k$is the corresponding interval length $1 \leq k \leq m * n$; (2) sort in descending order by area; (3) reading down the list of interval lengths, select each $L_k$where $L_k < L_{k+1}$. (4) Exclude intervals that include 0 or exclude the observed difference in medians. (5) Of the remaining intervals, the one with the largest area A_k is the optimal interval. See "Discussion" for comments on non-contiguous interval choices.

Figure 3.3. The four approaches to obtaining the desired interval. On the constructed posterior distribution (see figure 2), we pick an interval whose endpoints will be the confidence intervals on the median, and whose area under the curve is 1-p (where p is the desired p-value). When a one-tailed interval is desired (a), the longest interval is chosen that excludes 0, in the direction of the desired tail. When a two-tailed interval is desired, there are 3 main strategies. (b) finds the largest, maximally-compact interval that contains the observed median difference but excludes 0. Of the three two-tailed strategies, this one obtains the shortest interval. Strategy (c) finds the largest symmetric interval by length (L_1=L_2) that includes the observed median difference and excludes 0. Strategy (d) finds the largest symmetric interval by area (A_1=A_2) that includes the observed median difference and excludes 0.


a
One-tailed
$\mathcal{A}$
Observed median difference
0
$\mathcal{L} = \infty$
b
Two-tailed, min $\mathcal{L}_1 + \mathcal{L}_2$ with max $\mathcal{A}_1 + \mathcal{A}_2$
$\mathcal{A}_1$
$\mathcal{A}_2$
Observed median difference
0
$\mathcal{L}_1$
$\mathcal{L}_2$
c
Two-tailed, symmetric by length: $\mathcal{L}_1 = \mathcal{L}_2$
$\mathcal{A}_1$
$\mathcal{A}_2$
Observed median difference
0
$\mathcal{L}_1$
$\mathcal{L}_2$
d
Two-tailed, symmetric by area: $\mathcal{A}_1 = \mathcal{A}_2$
$\mathcal{A}_1$
$\mathcal{A}_2$
Observed median difference
0
$\mathcal{L}_1$
$\mathcal{L}_2$

Computation

The Bayes Box test, as well as data simulations, were written in MATLAB 2014a. Analyses were conducted using the Chronux toolbox (Bokil et al., 2010), modified Chronux codes, and scripts written in-house for importing and formatting EEG data.

Validation

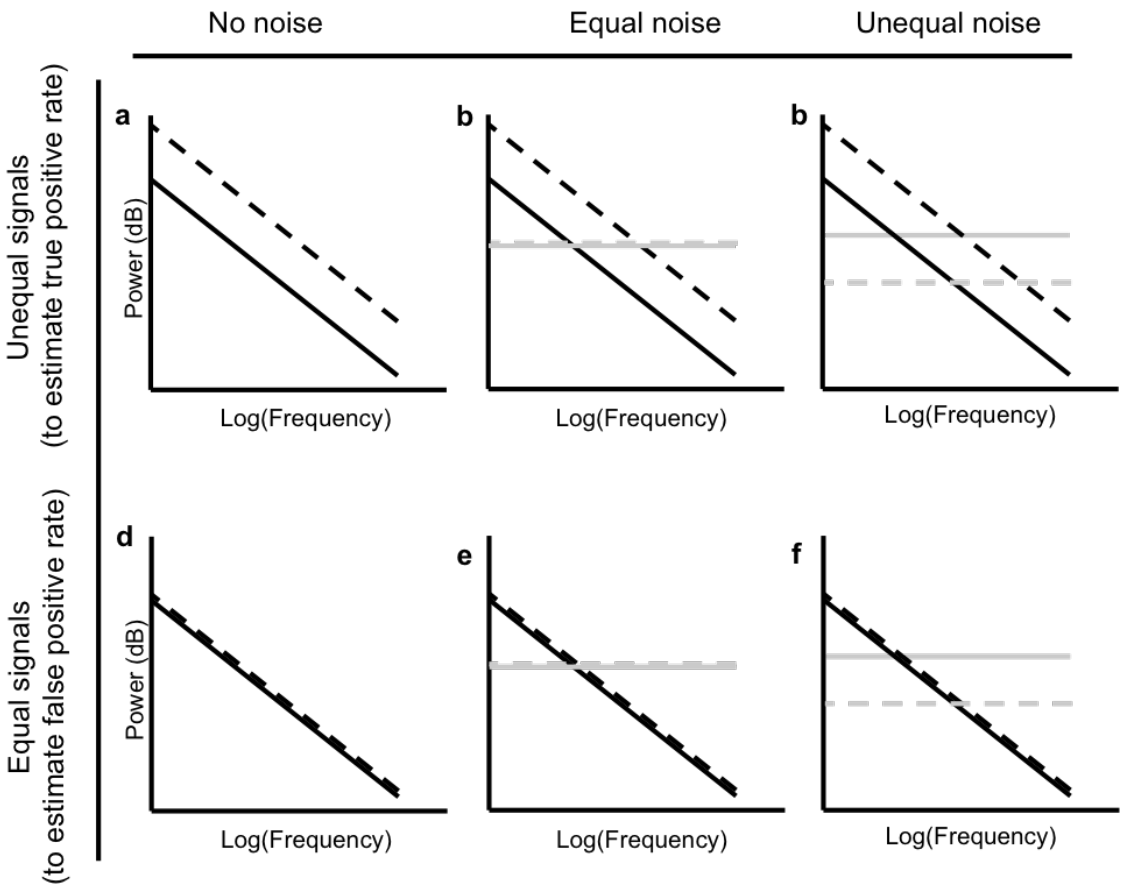
Simulated data

As in (Melman and Victor, 2016), simulated datasets were created by constructing a signal with one power spectrum, interspersed with large intermittent noise with another power spectrum. The signal, intended to simulate neural signal, was simulated with a $1/\omega$ power spectrum, where ω is the frequency. At each frequency, real and imaginary parts of Fourier components were drawn from a Gaussian with a variance of $1/\omega$; these components were inverse-transformed to obtain a time-domain signal. The noise, intended to simulate broadband artifact such as EMG, was modeled as bursts of large-amplitude white noise, introduced at poisson-distributed intervals. This was done by drawing the real and imaginary parts of the Fourier components from Gaussians with the same variance at every ωbefore applying the inverse transform. Data were simulated to have 30 segments with a length of 0.2 seconds each sampled at 250 Hz, and bursts of noise of 0.1 second

duration appearing in 0.2 of the segments on average. Noise bursts were drawn to have 3.3 times the power as the lowest frequency band of the signal, per unit bandwidth.

To construct ROC curves, we compared the performances of the five methods in three pairs of data sets. Each pair compares the detection of true positives from a pair of power spectra with the same 1/f power spectral shape but with an overall difference in amplitude (the higher-amplitude signal contained on average 1.58-times the power of the lower-amplitude signal), with the detection of false positives from a pair of power spectra with the same shape and same amplitude. The dataset pairs in comparison (1) consists of clean signals (true positives: **figure 3.4a**; false positives: **figure 3.4d**). The dataset pairs in comparison (2) contain poisson-distributed bursts of white noise noise spectrum that is equal between spectra being compared, and dominates at high frequencies (true positives: **figure 3.4b**; false positives: **figure 3.4e**). (3) Two signals whose noise spectrum is unequal and dominates at high frequencies (false positives: **figure 3.4c**); in the false negatives scenario (**figure 3.4f**) the higher noise is on the power spectrum with the lower *a* value.

Figure 3.4. Illustrations of power spectral shapes of simulated data. All are shown on a log-log scale. Black 1/f power spectra represent "signal", the component of the recording that is continuous. Gray flat white-noise power spectra represent "noise", the component of the recording that is intermittent. Noise was simulated as poisson-distributed bursts of large-amplitude white noise present in 20% of the segments on average. In each panel, two power spectra are shown, for two different signals to be compared: one solid and one dashed. Panels (a), (b), and (c) illustrate comparisons of two spectra with the same 1/f shape but a difference in overall amplitude; a statistical comparison is expected to produce true positives. Panel (a) illustrates two signals with an amplitude difference and no noise. Panel (b) illustrates the signals from panel (a) but with equal noise added to each signal: the gray lines represent power spectra of bursts of broadband artifact similar to EMG noise, simulated here with a white noise spectrum, present in several of the simulated data segments. The solid gray line is the power spectrum of the noise on the signal with the black solid line power spectrum; the dashed gray line is the noise on the dashed black line power spectrum. In this case, both spectra contain equal noise artifact. Panel (c) illustrates the power spectra of two 1/f signals containing bursts of white noise signals from two unequal power spectral distributions. The lower-amplitude signal (solid black line) contains the higher noise (gray solid line). Panels (d), (e), and (f) illustrate scenarios where the 1/f power spectra of the simulated neural signals have equal amplitude; a comparison test is therefore expected to produce false negatives. Scenarios are shown with (d) no noise, (e), equal noise, and (f) unequal noise.


No noise
Equal noise
Unequal noise
Unequal signals
(to estimate true positive rate)
Equal signals
(to estimate false positive rate)
a
b
b
d
e
f
Power (dB)
Log(Frequency)

Human data

We compared the standard and robust approaches' performance using two pairs of datasets from healthy subjects:

Comparison (1) A scenario in which artifacts are expected to lead to false positives. The two datasets consisted of epochs drawn from the same recording session of a healthy subject (labeled HC1) with eyes open. One dataset contained artifacts and the other did not; and

Comparison (2) A scenario in which the artifacts are expected to lead to false negatives and errors of sign. One dataset consisted of recordings from a healthy subject (labeled HC2) with eyes open, the other consisted of recordings with eyes closed.. In this case, two sets of epochs were drawn from each dataset: one that contained epochs contaminated by artifact, and one without.

Data were recorded using an FS128 head box and an augmented 10/20 montage. Input impedances were $\leq$5 KΩ. Data were sampled at 250 Hz and detrended. Data were visualized in Natus Neuroworks and imported to MATLAB using Eeglab.

Non-neural artifacts in the recording appear as outliers in statistical analysis. To create datasets with a controlled level of outliers, we controlled the amount of artifacts. First we identified artifact-free and artifact-containing three-second epochs by visual inspection. Artifacts included EMG, sweat artifact, and eye blinks (datasets (1) only). We then created "clean" and "artifact-containing" subsets by selecting

randomly from these epochs. For the datasets in comparison (1), the "clean" subset contained 30 epochs, and the "artifact-containing" subset contained 21 different clean epochs and 9 artifact-containing epochs.

For the datasets in comparison (2), each of the eyes-open subsets contained 17 epochs; the artifact-containing subset contained 12 different artifact-free epochs and 5 with artifact. The eyes-closed subsets each contained 10 epochs; in the artifact-containing subset, there were 7 different artifact-free epochs and 3 with artifact.

Data from HC1 was from the same recording used in [Melman Victor 2016], with a different selection of epochs.

Human subject participation was approved by the Institutional Review Board, and was consistent with the Declaration of Helsinki.

Results

We begin by comparing several significance tests on simulated datasets: two parametric methods (the standard two-group test, and a t-test on the means of the spectral estimates), and three non-parametric methods. This evaluation, based on simulated data, shows that the non-parametric methods can substantially reduce the level of false positives, false negatives, and errors of sign caused by outliers. We then

apply these methods to human EEG recordings, and the results show that these advantages hold for real data as well.

Simulated data

Histograms and performance correlation

In order to compare the tests' sensitivity, spectral comparisons under three noise scenarios were carried out for pairs of spectra with a true difference in overall amplitude (diagrammed in **Figure 3.4** panels **a, b,** and **c**). The three scenarios consist of: (1) no artifact (**figure 3.4a**), (2) equal levels of artifact in both spectra (**figure 3.4b**), and (3) a higher level of artifact in the lower-power spectrum (**figure 3.4c**). **Figure 3.5** summarizes results from all 5 tests as histograms of P-values obtained over 100 data simulations. Results are from one-tailed implementations of the test. A high frequency band was chosen for comparison, where the broadband noise (when present) dominates the 1/f spectrum of the underlying signal. The histograms show that all 5 tests are sensitive when artifact is absent, but diverge in performance when artifact is present. The parametric methods have an increased false negative rate (column 2) and make errors of sign (column 3) -- as expected since in these scenarios, the artifact dominates the power in the underlying signal. Both kinds of errors are less frequent with the non-parametric methods.

Figure 3.5. Histograms of p-values from applying the one-tailed version of the 5 comparison tests on 3 simulated data scenarios over 100 simulations. Comparisons were done at one frequency (light gray dotted vertical lines on power spectral diagram); a high frequency was chosen to illustrate cases where noise dominates. Noise is present in about 20% of the data segments, with 30 segments in each data set. Column 1 shows results from comparing two clean signals separated by a difference in powers (see figure 4a). Column 2 shows results from comparing signals separated by a difference in powers with equal noise (see figure 4b). Column 3 shows results from comparing signals separated by a different in powers with unequal noise, with the larger-amplitude noise on the smaller-amplitude signal (see figure 4c). Results show that the two parametric test—the t-test and the two-group test—are more sensitive to noise, finding false negatives when noise dominates (column 2), and produces errors of sign when the difference in noise spectral power is in the opposite direction of the difference in signal spectral power (column 3). The robust methods—the rank-sum, permutation test, and Bayes box—all exhibit relative insensitivity to large intermittent outliers.

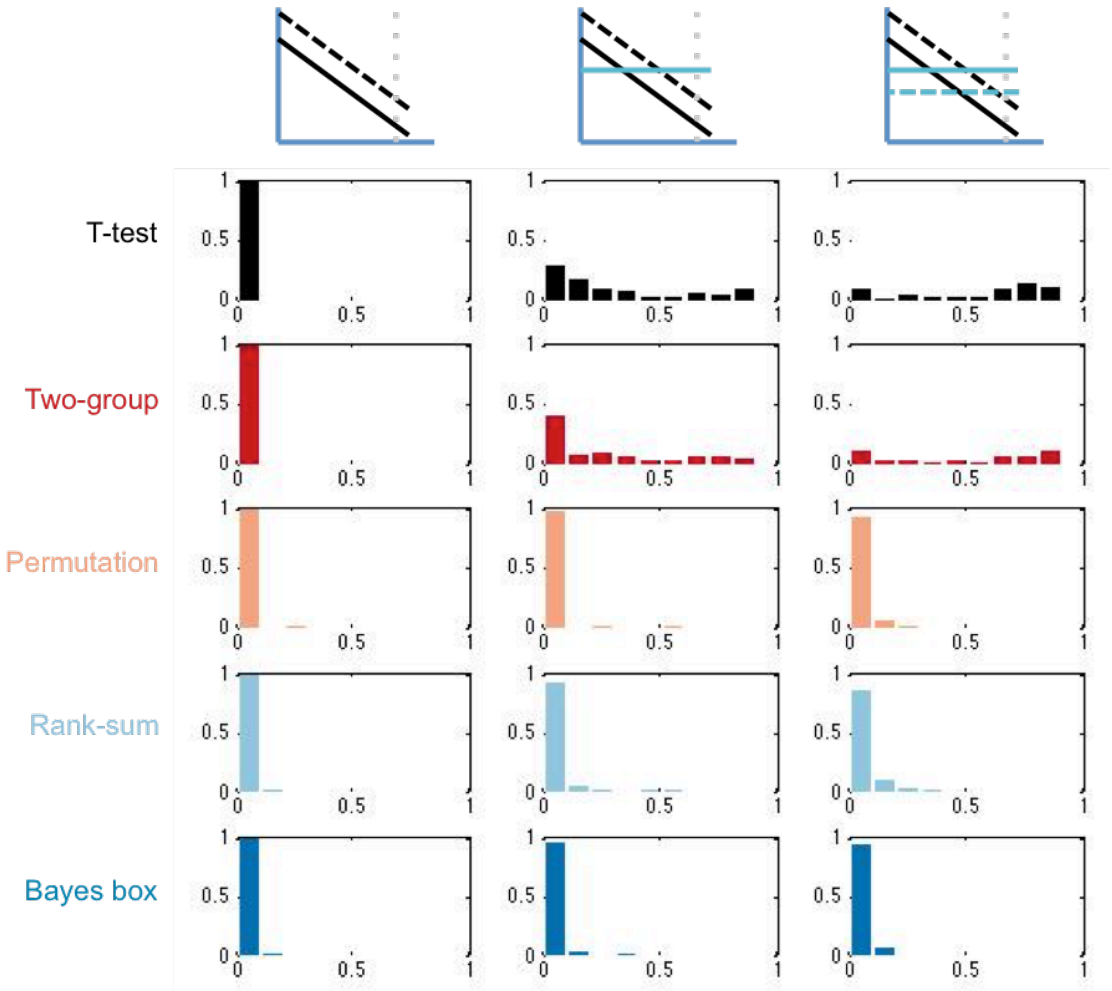

T-test
Two-group
Permutation
Rank-sum
Bayes box
0
0.5
1

Figure 3.6 highlights how the behavior of these methods cluster into two groups, based on how p-values from individual runs of the simulated data are correlated between the 5 methods. Results from the three non-parametric/robust methods are highly correlated, and distinct from the parametric/non-robust methods. Within the non-parametric/robust groups, the two quantile-based tests--the median permutation and the Bayes box-- are more tightly correlated with each other than with the rank-sum test under noise-dominated scenarios.

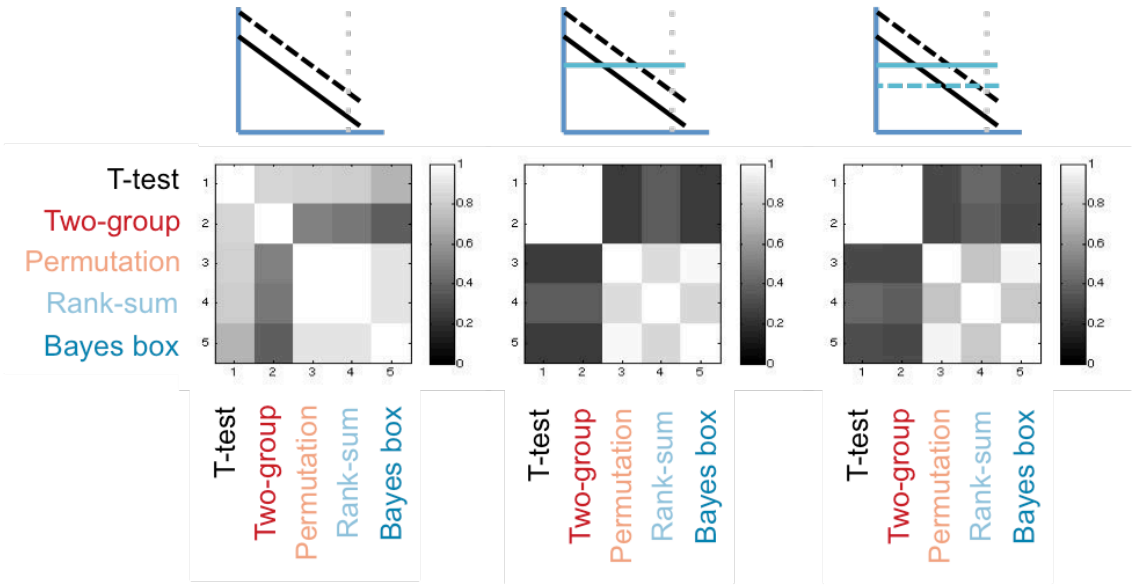


Figure 3.6. Correlations of p-value results of 5 comparison tests on the same simulated data sets (over 100 simulations). Results from figure 3.5 were correlated and are shown here from the 3 spectral comparison scenarios: two clean signals, two signals containing equal noise, and two signals containing unequal noise, with a difference opposite the direction of the difference in signals. Correlations are highest on clean data: all the tests correctly identify true positives. On noisy data, the two parametric tests cluster most closely to one another and the three nonparametric tests cluster to each other. The rank-sum clusters less strongly with the median-permutation and Bayes box, reflecting its slightly higher sensitivity to differences in data distribution.

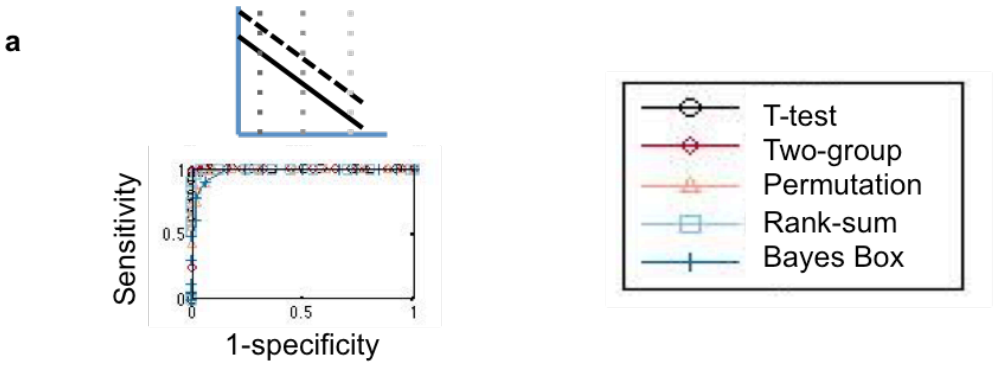
ROC analysis

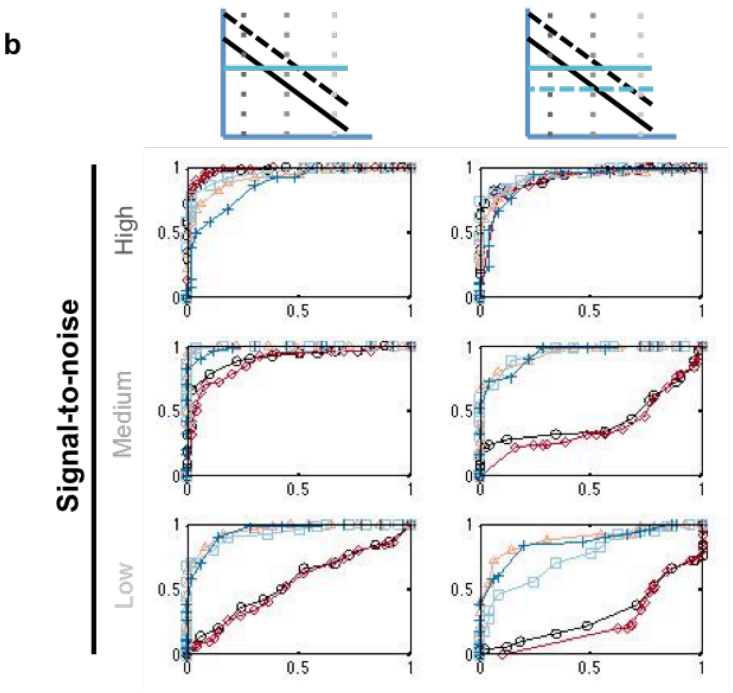
To determine whether the reduction in false-negatives by the non-parametric methods incurred an increase in false positives, we then constructed receiver operating characteristic (ROC) plots for the 5 tests under the three different noise scenarios. ROC curves plot the true positive rates (sensitivity) on unequal signals (**figure 3.4**, panels **a**, **b**, and **c**) vs. false positive rates (1-spcificity) on equal signals (**figure 3.4**, panels **d**, **e**, and **f**). **Figure 3.7a** shows results from clean signals (**figure 3.4a** and **3.4d**) . Behavior is similar at all frequencies when there is no artifact (the spectra differ by a constant factor), so only one frequency is shown. **Figure 3.7b** plots the ROC curves when artifact is present: equal levels in column 1 (**figure 3.4b** and **3.4e**), and unequal levels in column 2, with the greater artifact on the lower-amplitude signal (**figure 3.4c** and **3.4f**). For both noise-containing scenarios, signal dominates at low frequencies, so we expect fewer false negatives or errors of sign; at high frequencies, artifact dominates, so we expect more false negatives or errors of sign. **Figure 3.7c** tabulates the area under the ROC curves (AUC's).

Figure 3.7 shows that on artifact-free data (**figure 3.7a**), the area under the curve (AUC) of the ROCs all approach 1, with the t-test and the two-group test performing slightly better than the three non-parametric tests due to their greater efficiency in cases with no outliers. In **figure 3.7b**, ROC curves are plotted for scenarios when artifact is present. As the level of artifact increases (here, corresponding to higher frequencies) both the t-test and the two-group test performed progressively worse, while the parametric tests were all relatively unaffected. In the scenario in which both groups have the same level of artifact, the t-test and two-group

tests yielded an AUC of ~0.5 for the highest artifact level, while all three parametric tests yielded AUC's well above 0.7. In the scenario in which the group with the lower-amplitude spectrum had a greater amount of artifact, the parametric tests yielded an AUC's less than 0.5, indicating that they confounded the artifact with the underlying EEG signal, and made errors of sign. In contrast, the ROC curves for all three non-parametric tests remained above the diagonal—they continued to find more true positives than false positives, with the proper sign of the comparison, even in the presence of confounding artifact.

Figure 3.7. ROC curves on simulated data. Panel (a) shows ROC curves on clean data. All 5 methods achieve good results, with the areas under the curves approaching 1; the Bayes box test and permutation tests have slightly smaller areas under the curve because of their lower efficiency. Panel (b) shows results from the equal-noise and unequal-noise scenarios, the first to illustrate false negatives and the second to illustrate errors of sign. ROC curves from spectral comparisons at three different frequencies are shown: since signal dominates at low frequencies and noise dominates at high frequencies, this illustrates the methods' performance on varying signal-to-noise ratios. Noise was present in about 20% of segments. ROC curves and corresponding areas under the curves show that the Bayes and permutation tests perform worse when signal-to-noise ratio is high due to lower efficiency; however, when the signal-to-noise ratio is low, the permutation and Bayes box demonstrate much better performance in recovering both true positives and correct direction of difference. Rank-sum performs well under most of the cases, but exhibits somewhat more sensitivity to outliers, especially in the unequal-noise low signal:noise scenario.





Areas under curves:

High	.98	.93
	.98	.90
	.92	.92
	.96	.95
	.86	.90
Medium	.91	.44
	.88	.38
	1.0	.94
	1.0	.95
	.98	.93
Low	.56	.31
	.53	.25
	.96	.89
	.93	.75
	.94	.86

Human data

The rank-sum, median permutation, and Bayes box tests were all run on the human data and performed similarly to each other relative to the parametric methods, with a few considerations:; however, since the rank-sum was slightly less robust in noise-dominated scenarios than the median permutation or Bayes box tests, and the Bayes box test showed less efficiency than the other non-parametric methods. Since the amount of clean data in the healthy control recordings was limited, we opted to test the more median permutation test against the parametric two-group test on human data for its efficiency and robustness.

The first comparison (HC1) was designed to test whether the robust methods would reduce the level of false positives. To address this, we compared spectra from two datasets consisting of epochs from the same eyes-open recording: one consisting of epochs hand-selected to be artifact-free, and the other consisting of epochs in which 70% were artifact-free and 30% contained eye movement and EMG artifacts (see Methods for details). **Figure 3.8** shows spectral estimates from two representative channels (FP1-F3 and FP2-F8). The standard (two-group test) finds a statistically significant difference at many frequencies in the range of 1-15 Hz in both channels, and at higher frequencies in FP1-F3. In contrast, the robust method (median permutation test) estimate of difference of spectra finds statistically significant differences in only a few frequency bins around 10 Hz for FP1-F3 and below 10 Hz for FP2-F8.

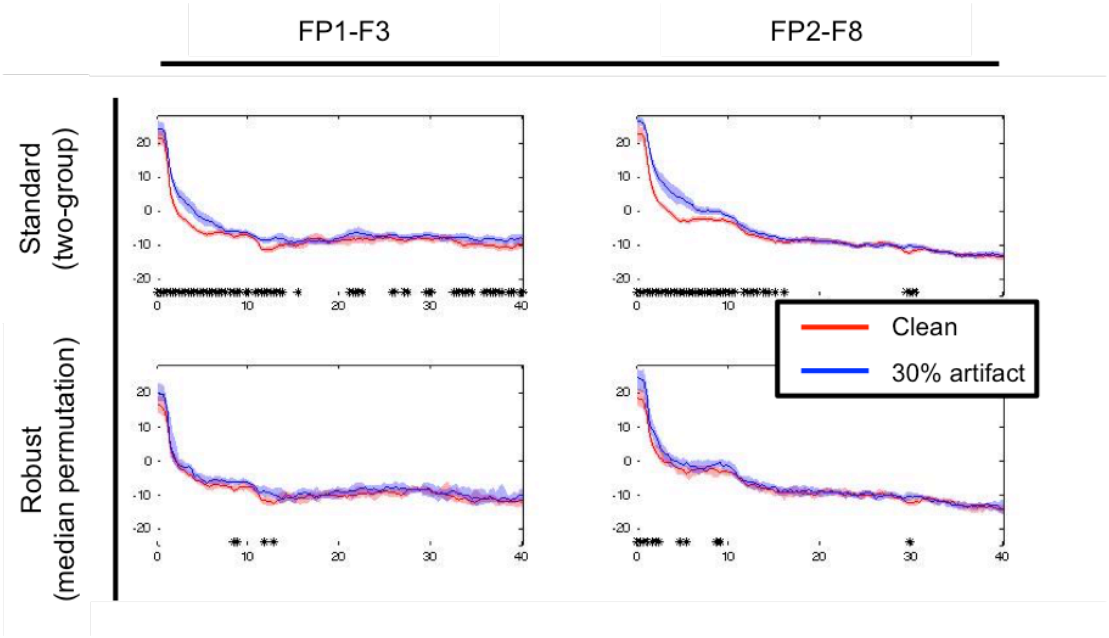


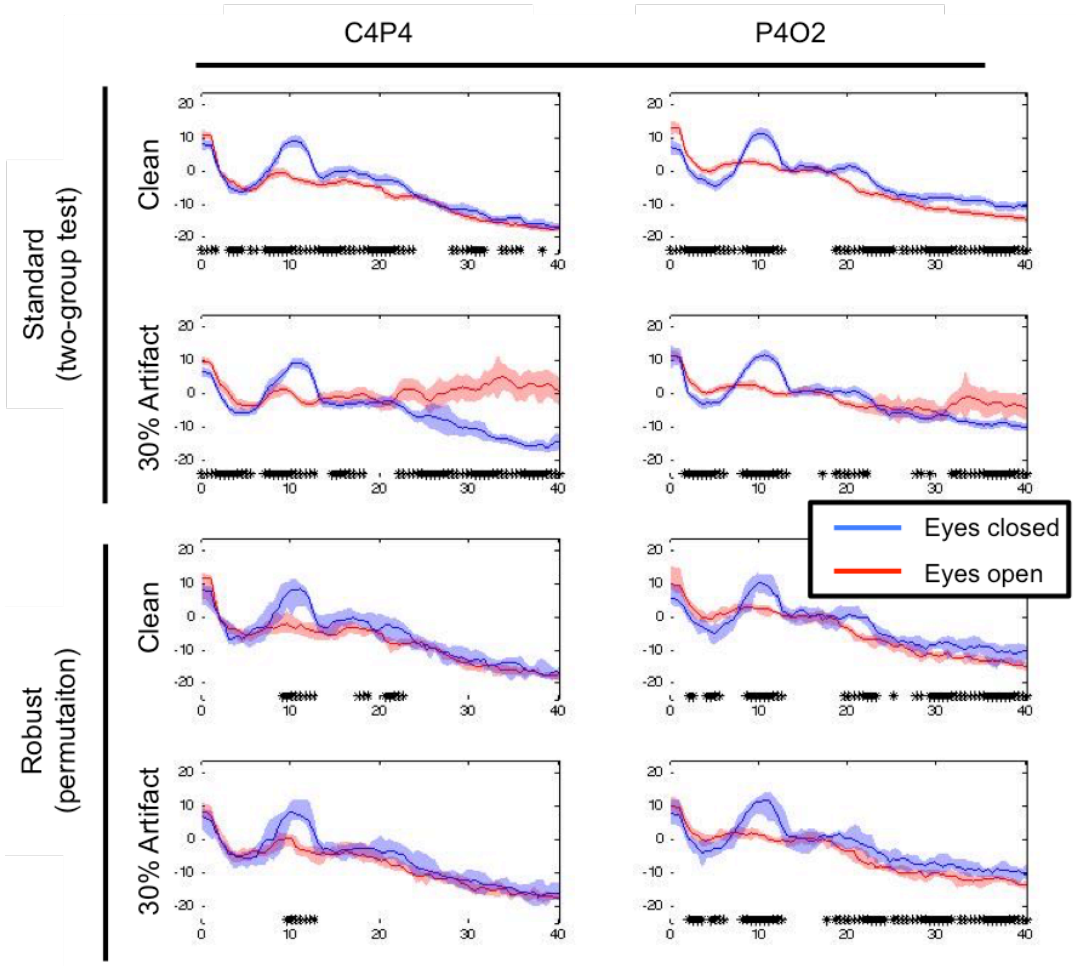
Figure 3.8. Power spectra from a healthy control (HC1) baseline recording under awake resting state conditions. Clean (red) and noisy (blue) power spectra are shown from two frontal channels using the standard (two-group-test) and robust (median permutation test) approaches. Black asterisks underneath each spectrum indicate frequency bands found to be significant ($p<0.05$) after FDR correction (with FDR = 0.05). With the robust comparison, fewer frequency bands were marked as significant, demonstrating a reduction in false positives due to differences in noise.

The second comparison (from HC2) was designed to be a scenario in which artifact could lead to false negatives and errors of sign. The two spectra compared were EEG recordings from two different states: awake with eyes open, and awake with eyes closed. Significance testing was done on two subsets of epochs: one that was artifact-free, and one in which 30% of the segments had artifacts. We anticipate seeing a difference in power in the 8 to 12 Hz band in the posterior leads under these conditions, based on the known increased prominence of the alpha rhythm with eyes closed is known to produce an alpha peak originating in the occipital lobe and detected in posterior channels in healthy controls (Britton, 2016; St. Louis et al., 2016). We also expected that the composition of non-neural artifact levels will differ across the two datasets -- for example, the eyes-open condition will contain eye blink artifacts, which are absent in the eyes-closed scenario.

Figure 3.9 shows the results of the analysis. With regard to the alpha peak, the expected change is detected in both C4-P4 and P4-O2, whether or not artifact is present, and with either analysis method. In other frequency bands, the analysis methods yield different results. Above 20 Hz, the higher level of artifact in the eyes-open condition leads to an error in the sign of the comparison for the standard method, in both channels: the eyes-open condition appears to have greater power. With the robust method, the sign of comparison is correctly registered in P4-O2; in C4-P4, where the difference in spectra is less-prominent, it fails to detect a difference. A similar phenomenon is seen at the low-frequency (1-7 Hz) range : in the absence of non-neural artifact, the eye-open dataset has a slightly greater power; the standard method makes an error in sign for P4-O2, while the robust method identifies the

correct sign in this channel, and fails to detect a difference in C4-P4. In sum, both methods detect the substantial and expected changes in alpha-band power across states. With regard to more subtle changes, in frequency ranges in which the standard approach makes errors of sign, the robust approach either identifies the correct sign of the comparison, or fails to find a significant difference.

Figure 3.9. Power spectra from a healthy control (HC1) baseline recording under two different awake resting state conditions: an eyes open baseline and eyes closed baseline, from one central and one posterior channel. Results are shown from comparing clean data and 30% artifact-containing data from each state, using the standard (two-group test) and robust (median permutation test) approaches. Results show that while artifact introduces false positives in C4P4, especially at higher frequencies, the robust method is not affected. Similarly, in P4O2 the standard method finds a reversal in direction of significance at high frequencies due to noise, while the robust method preserves the direction of significance. The robust method also retains the ability to detect significance in the 10-12 Hz alpha band, where a power increase in the eyes-closed condition is expected.


C4P4
P4O2
Standard
(two-group test)
Robust
(permutaiton)
Clean
30% Artifact
Clean
30% Artifact
Eyes closed
Eyes open

Discussion

In the case of comparing two power spectra, a small number of large outliers can skew the estimated significance. Since many non-neural signals in EEG are large in magnitude but intermittent, the effect of these outliers is particularly problematic for EEG data analysis. We showed that applying a non-parametric method to both simulated and recorded datasets mitigates somewhat the effect of these outliers on the spectral comparison. This is the case when outliers create both false positives as well as false negatives, which distinguishes the non-parametric approach from multiple correction adjustments such as the false discovery rate correction (FDR) (Yoav Benjamini, 1995), which only reduce false positives but do not solve problems with false negatives or errors of sign. We also show that quantile-based tests such as the permutation test of medians and the Bayes box test show slightly more robustness under the noise scenarios presented here than the rank-sum test, which is sometimes used in EEG across-subject frequency comparison (Lee et al., 2018).

Another advantage of quantile-based approaches is that they provide confidence intervals on the difference in quantiles/medians. The Bayes box test calculates confidence intervals explicitly; the permutation test can estimate confidence intervals using the bootstrap method. The rank-sum can give confidence intervals using various approaches (Bauer, 1972; Geyer, 2003) not discussed here, but care has to be taken to prevent implicit assumption of paired samples, equal

variances, or symmetric distributions. The Bayes box and permutation test make no such assumptions.

To reiterate the need for the Bayesian approach, it allows for the estimation of confidence intervals on the non-parametric median statistic on an undefined distribution. We assume a flat prior distribution in the space of order statistics; this is integrable (in the space of order statistics), and leads to a conservative estimate of the confidence interval. Note that the construction of the posterior presented here assumes statistical independence between two sets of samples; this simplifying assumption holds under many cases of EEG datasets; they may be accounted for in future iterations of the test, but not without re-introducing assumptions about the distribution of the data. One last point of discussion on this Bayesian approach is the choice here of contiguous intervals rather than the potentially non-contiguous maximum a posteriori probability (MAP): although the possibility exists of implementing the MAP to retain distributional information, here we are concerned with obtaining easily interpretable confidence intervals for our parameter estimate, and therefore contiguous intervals are more appropriate.

One drawback to the robust approaches is the low efficiency of the methods relative to parametric approaches. This is an expected tradeoff between robust and non-robust estimators; therefore the amount of data available is a bigger consideration with the robust approach, particularly the Bayes box test. However, this is partially mitigated by the greater tolerance of the robust approach for artifact-containing data.

In conclusion, quantile-based approaches to estimating the differences between power spectra are robust in the presence of large outliers typical of EEG data, and can be easily incorporated into standard analysis pipelines. Although they do not replace standard outlier handling methods in many cases, they can be implemented in conjunction with the existing toolbox.

Acknowledgments

Credit goes to Dr. Mary Conte for collection of EEG healthy control baselines, and to Dr. Andrew Goldfine for many of the scripts for data collection and processing, including the basis for our permutation test code.

CHAPTER 4: ZOLPIDEM - A CASE STUDY

In this chapter, I apply the robust approach to a dataset of baseline EEG recordings from a study demonstrating the paradoxical activating effect of the GABA-A agonist Zolpidem on a select cohort of minimally conscious state (MCS) patients. Baseline recordings present a potentially useful application of median-based spectral estimation since much more data are available than for ERP and other short recordings, and hand-picking segments in these datasets is particularly tedious.

For diagnosis and assessment of DOCs, baseline EEG can be used to probe brain network integrity, an important diagnostic marker of latent brain function (Bai et al., 2017). An alternative EEG paradigm is measuring event-related potentials (ERPs), where EEG responses to external stimuli--such as auditory or visual cues--are used to indicate cognition. The drawback of ERP analysis is it depends on intact perceptual pathways as well as adequate patient arousal and cooperation, which are commonly affected following an injury. In contrast, baseline resting-state EEG recordings capture spontaneous network activity, which can reflect latent network dynamics. Because different states of consciousness have different characteristic baseline EEG power spectral features (Forgacs et al., 2017; Schiff et al., 2014), observing temporal changes to the power spectrum can reveal temporal state change dynamics. The spectrogram is a tool which compiles multiple individual spectra into a single image in order to visualize temporal dynamics of frequency features over the timecourse of Zolpidem's washout.

Zolpidem is one of a handful of drugs that have a paradoxical activating effect in a handful of patients in minimally conscious state (MCS) (Thonnard et al., 2014). Though typically prescribed for insomnia, as it induces drowsiness, Zolpidem has been reported to produce a paradoxical increase in awareness. In patients with DOCs, this paradoxical activation can manifest in increased awareness of or reactivity to external stimuli. Responsiveness is rare (estimated at 6.7% (Thonnard et al., 2014; Whyte and Myers, 2009)) and dependent on the mechanism of injury, type of brain damage, and subjet's individual neuroanatomy; however, the improvement in in behavioral markers of consciousness, as measured by the JFK Coma Recovery Scale--Revised (CRS-R) or other behavioral scoring paradigm, can be drastic. CRS-R improvement is accompanied by a concurrent shift in EEG spectral features toward a shape more resembling normal awake EEG (Forgacs et al., 2017; Schiff et al., 2014).

Interpretation of EEG spectral features in zolpidem responder studies

Several studies have found similar EEG correlates of post-zolpidem improvement (Thonnard et al., 2014; Whyte and Myers, 2009; Williams et al., 2013). In (Williams et al., 2013), three subjects who exhibited responsiveness to zolpidem both behaviorally and electroencephalographically were reported. The main finding was a consistent reduction in what they termed an abnormal alpha peak in the EEG power spectrum centered around 7 Hz, as well as an increase in 15-30 Hz broad beta-band activity, from before the dose to after, with features that persist post-washout. In (Carboncini et al., 2014), similar spectral shifts were shown in a midazolam-

responsive patient, where an abnormal 7 Hz peak (termed the theta2 band) attenuated and 15-30 Hz activity increased. Although (Williams et al., 2013) declined to speculate that the 7 Hz peak was predictive of zolpidem responsiveness, (Carboncini et al., 2014) speculated that the peak was predictive not necessarily of zolpidem responsiveness, but of responsiveness to GABA-A agonists in general. (Williams et al., 2013) instead suggested that the 7 Hz rhythm is the intrinsic firing rate of cortical cell membranes (Silva et al., 1991), indicating a deafferentated but intact cortical network but no potential to predict drug response.

Resolving whether EEG dynamics are due to Zolpidem's effect on corticothalamic or corticocortical connectivity changes has been addressed by (Williams et al., 2013), who argues that thalamic involvement is likely since a reduced beta peak has been speculated to arise from widespread thalamic deafferentation. Restoration of this peak in responders after zolpidem treatment co-occurs with an increase in JFK coma recovery scale-revised (CRS-R) score, likely from restoration of thalamic activity. The restored beta and concurrently reduced theta activities revert back to baseline over the course of about 3 hours, and abnormal theta returns.

To visualize the temporal changes in EEG over a long period of time, a spectrogram is used. The spectrogram is a matrix which can be visualized as a heatmap, where each column is a consecutive power spectra from a short period of time. Here, I focus on the low- to mid- frequency bands (1-45 Hz) region of the spectrogram, to parallel the analysis done in (Williams et al., 2013). Here we compare the standard and robust approach to spectrograms on these data to determine the

differences between the two and in which cases one might be preferable to the other. We also use the tool to observe and analyze spectral features of significance for the understanding of the process of recovery in responders under zolpidem.

Methods

Patient data collection

EEGs were recorded from patients admitted to New York Presbyterian's inpatient neurology service over the course of several days. On top of their standard at-home medication regimen, zolpidem was administered in 10mg doses either through oral syringe or percutaneous endoscopic gastrostomy (PEG) tube. Video was recorded simultaneously. Behavioral assessments using the CRS-R were done throughout the day, both before and after zolpidem.

Subject IN301 (subject #2 in (Williams et al., 2013)) is a 32 year old male in MCS lasting for 2 years. A near-drowning in a motor vehicle accident 9 years prior to the study resulted in a combination of trauma and hypoxic-ischemic injury. Zolpidem responsiveness was first observed after 2 years. While off zolpidem, subject exhibits a resting tremor on his right side, prolonged latency in responses to stimuli, and loss of oral feeding capacity. Subject recovered speech and deglutition on zolpidem, with short latency of responses.

Subject IN376 (subject #1 in (Williams et al., 2013)) is a 45 year old male who suffered a TBI after falling off a ladder onto a concrete surface 5 years prior to the study and was categorized as MCS by a CRS-R assessment. While off zolpidem,

the patient was able to stand up and move from the bed to the wheelchair, but had only intermittent responsiveness to commands. On zolpidem, the patient could demonstrate use of several common objects (comb, toothbrush, spoon) and accurately give his name, write simple words, and answer some general questions.

Both subjects were medically stable with non-progressive injuries..

Statistical tools

Data were exported using Natus Neuroworks software. Analysis was done in MATLAB using Chronux (Bokil et al., 2010) and in-house codes for data processing and formatting. Median spectrograms were calculated using modifications of codes from (Melman and Victor, 2016).

Spectra

Power spectra were calculated from 1 hour of data before zolpidem was given, and from 1 hour after zolpidem. Datasets were segmented into 3-second epochs. All epochs were used with no artifact removal. 5 tapers and a time-bandwidth product were used. Standard spectra are shown with jackknife error bars; robust spectra are shown with Bayesian error bars.

Spectrograms

With the goal of comparing results to (Williams et al., 2013), who calculated one spectrum per minute by using the cleanest 3-second segment of each minute, I constructed one spectrum for each minute of the recording using all the data, with

four approaches. The first was to take spectra from unsegmented minutes of the recording, to create a raw spectrogram. The second involved segmenting each minute into 3-second segments and take the standard approach (mean over segments) to construct the mean-smoothed spectrogram. The third was similar to the second, substituting the robust approach (median over the segments) for the standard approach to obtain the median-smoothed spectrogram. 5 tapers were used with a time-bandwidth product of 3. These three spectrograms were adjusted for optimal viewing by flattening the top 5% and bottom 5% of values. The smoothed spectrograms—both the mean and the median—are similar to the coherogram smoothing performed in (Wong et al., 2011). The last spectrogram is the mean-median-difference spectrogram, where the values from the median spectrogram (in decibels) were subtracted from the mean spectrogram (in decibels).

Robust reanalysis results

Spectra (standard and robust)

The first exploration was to assess the effect that using the robust method over the standard method had on power spectra. **Figure 4.1** shows the power spectrum on an hour of data pre-dose and an hour after in responder IN301. Both the standard and robust power spectra revealed the 7.5 Hz abnormal alpha peak reduction and return of normal beta centered around 25 Hz reported in the original paper. The methods diverged in performance in a few key ways: first, in both the pre- and post- zolpidem spectra, delta and theta powers are lower with the robust method than with the

standard method. Overall power, in particular at high frequencies, was reduced, indicating a reduced contribution from broadband artifact.

The standard method identified a feature that was not seen by the robust method: a 10 Hz peak that occurred with, but was distinct from, the abnormal 7.5 Hz alpha, which is picked up by the standard method but not by the robust method. This feature appears in both the pre- and post-zolpidem conditions, though it is attenuated somewhat following zolpidem administration. The loss of this peak motivated spectrographic analysis in order to better understand the dynamics of this peak.

Power spectra from IN376 differed slightly (**figure 4.2**). Here, both the mean and the median showed an alpha peak at 10 Hz. The robust method better detected the distinct 7 Hz and 10 Hz rhythms pre-dose, which are attenuated in the next hour after the dose.

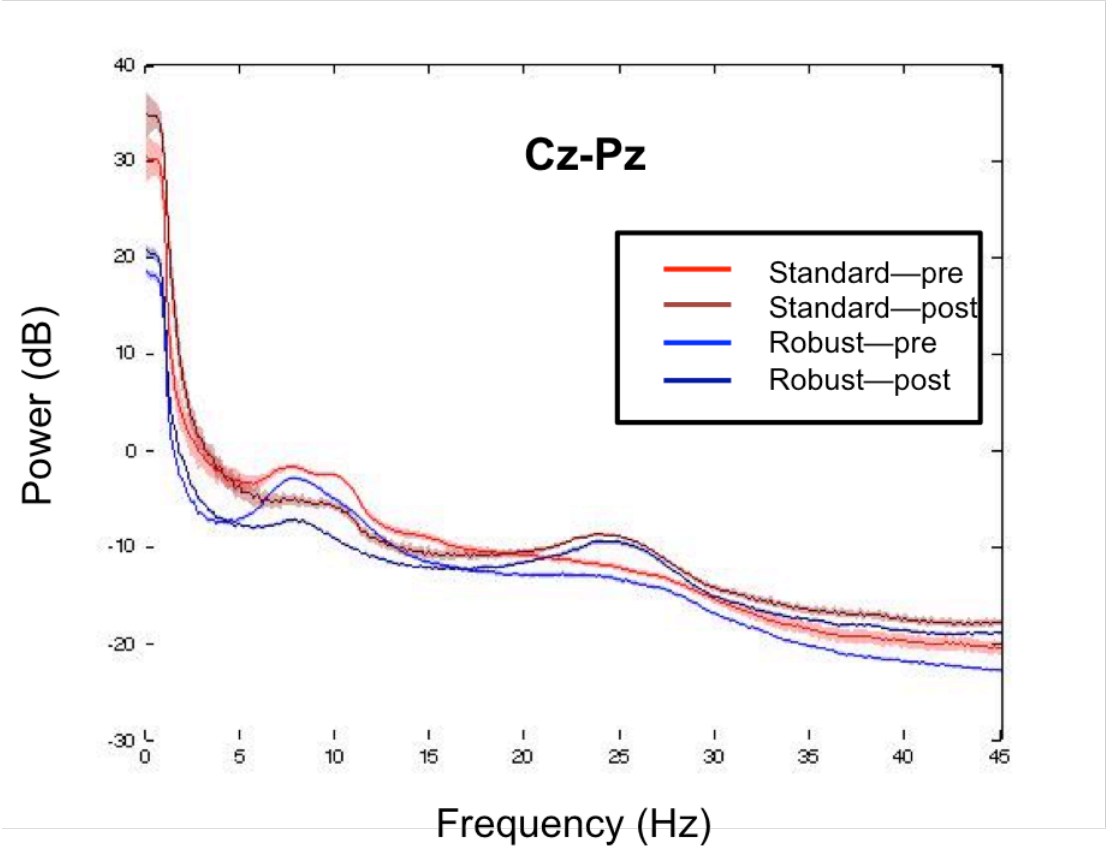


Figure 4.1. Power spectra from the first dose of day 2 in subject IN301 from 1 hour of data pre- and 1 hour post- zolpidem dose, with no artifact removal, from channel Cz-Pz. The use of the robust method greatly reduces low-frequency power and somewhat reduces the high frequency power; the robust method also loses the 10 Hz peak observable with the standard method.

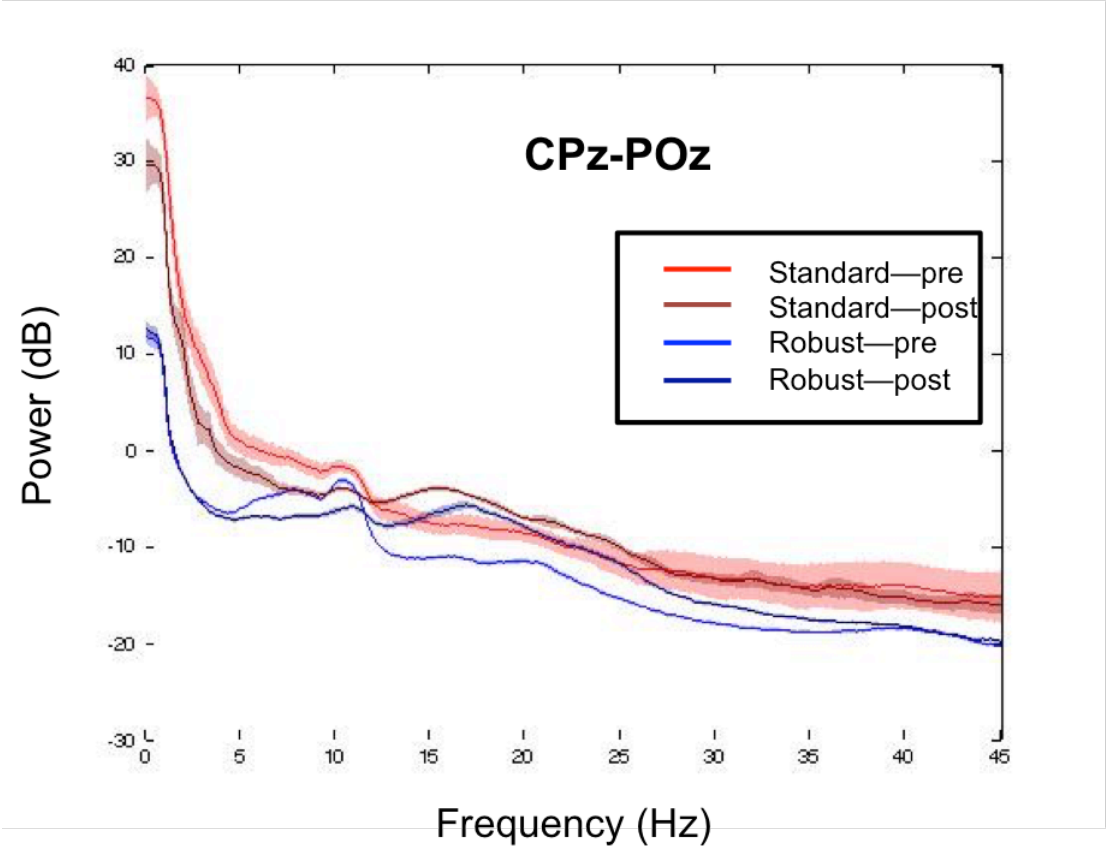


Figure 4.2. Power spectra from the first dose of day 3 in IN376. Here the pre-dose alpha peak is more distinct in the robust spectrum (light blue), and the standard pre-dose spectrum (light red) does not capture the 7 Hz peak due to low-frequency noise.

Spectrograms (regular, smoothed standard, smoothed robust)

To better understand the temporal dynamics of the 10 Hz peak that attenuates after zolpidem and was only obserable in the power spectrum using the standard method, I applied the robust method to spectrograms on a longer superset of the data used to calculate the power spectra. Results are shown in **figure 4.3**, which illustrates that both the mean- and median-smoothed spectrograms demonstrate a reduction in noise artifact and a cleaner background when compared to the raw spectrogram. The overall background in the robust method is lower, making the beta peak centered around 25 Hz more prominent. The difference between the two methods is most apparent in the 10 Hz band shown in the power spectra from **figure 4.1**: namely, that pre-dose and post-dose regions in the mean power spectrum clearly show two different activity bands: one at 7.5 Hz and one at 10 Hz. This feature does not appear in the robust power spectrogram. This is undoubtedly the result of finer-grained temporal dynamics: the 10 Hz activity appears intermittently and in fewer than half of the 20 epochs of 3-second duration in the recording. The bottom panel of the figure shows the difference between the mean and median spectra, which predominantly reveals intermittent features. The 10 Hz peak is the only feature, aside from some broadband movement and EMG artifact mostly concentrated in the 4-4.5 hour window, that have these temporal dynamics. **Figure 4.4**, spectrograms from IN376, does not capture similar dynamics; instead, the alpha peaks both before and after zolpidem are continuous. Further inspection of the signal (**figure 4.5**) with concurrent video (not shown) reveals that the alpha rhythm correlates with eye closing during recording, a characteristic of normal posterior alpha dynamics.

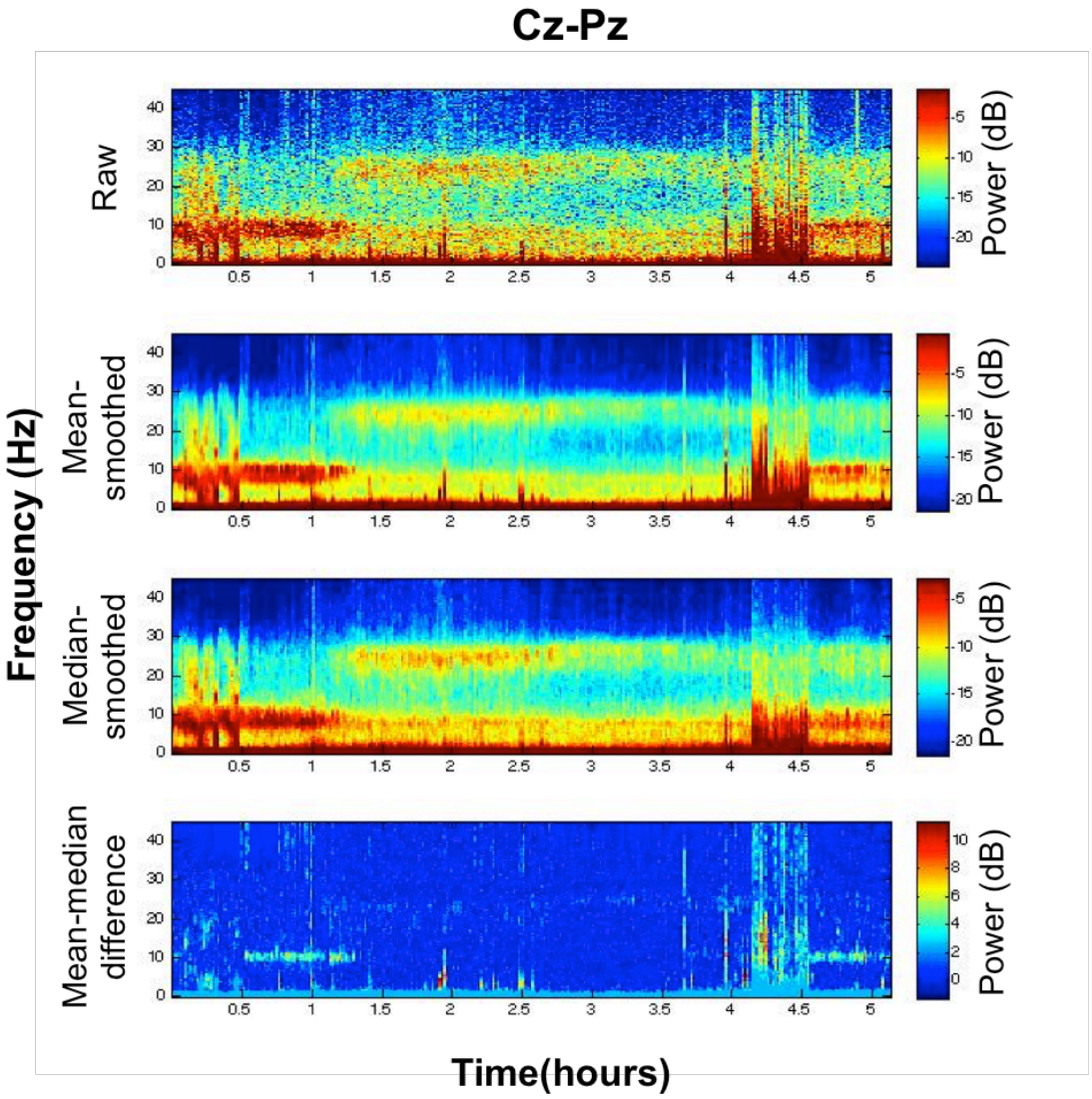


Figure 4.3. Spectrograms from the first dose of day 2 in IN301 over 5 hours before, during, and after drug dose. Drug was given at hour 1.

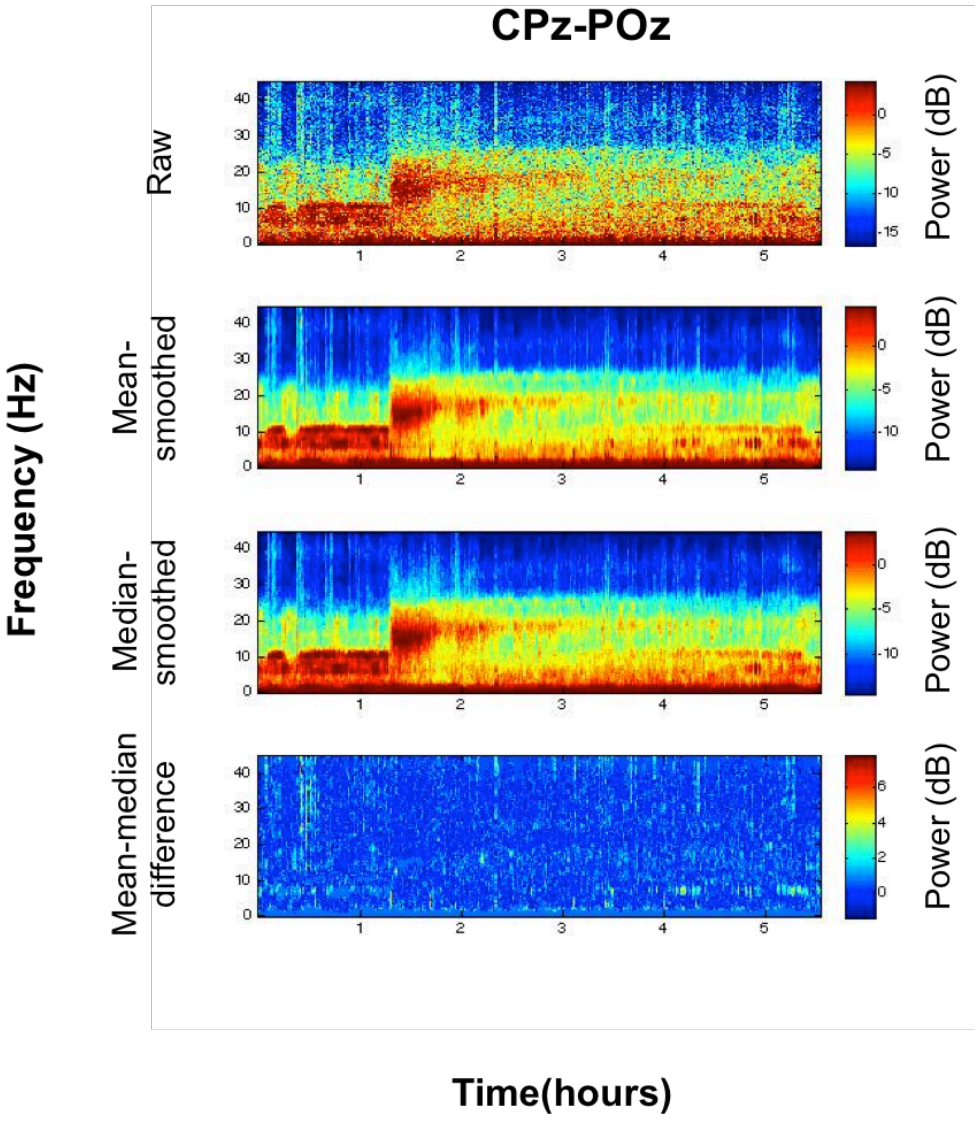


Figure 4.4. The power spectra from the first dose of day 3 in posterior channel CPz-POz from IN376. Drug was given at hour 1. Here the 10 Hz alpha peak is picked up by both the mean and the median, in both the pre-dose and post-washout states. Little difference is picked up in the mean-median spectrogram, indicating different dynamics than shown in figure 3.

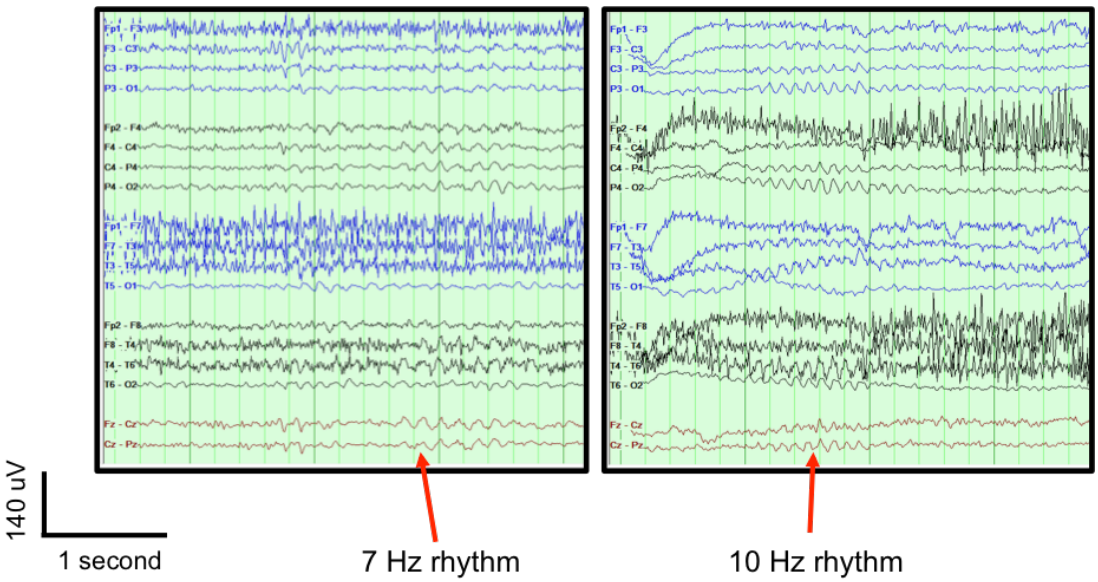


Figure 4.5. EEG recordings from IN301. On the left is a tracing includes the 7 Hz rhythm; on the right is a tracing that includes the 10 Hz rhythm. The 10 Hz rhythm appears during times of increased wakefulness, which tend to include more noise. This highlights the importance of using more or all of the data, rather than the cleanest segments. In this case, the patient had their eyes closed for the couple seconds, generating the alpha that appears in the posterior channels.

Discussion

In this case study, I applied the robust spectral approach to EEG spectra and spectrograms of minimally conscious patients who exhibit a paradoxical activating response to zolpidem. This demonstrated the ability of the robust method to not only utilize all data available, but also to better extract spectral features necessary for accurate assessment and diagnosis of disorders of consciousness using the EEG.

In the power spectra, the reduction of low-frequency (0 to 6 hz, including delta and theta bands) was significant when all the data were included in the analysis. Since low-frequency activity is inversely correlated with consciousness, and has particular discriminatory power between UWS and MCS (Lehembre et al., 2012; Sitt et al., 2014), a reduction in noise in this frequency range even when all data are used has potential for improving DOC state discrimination. It may be of use in further development of automated EEG classifiers.

This analysis demonstrated the ability of the both the median- and mean-smoothed spectrograms to obtain spectrograms with no hand-cleaning of the data before plotting. The median-smoothed spectrum removed some of the intermittent, broadband EMG artifacts, which produces a cleaner spectrum on which to apply further computational analyses. The features and temporal dynamics observed were consistent with the (Williams et al., 2013) analysis, showing that spectrograms capture several distinct EEG spectral features after the zolpidem dose: most notably, a 15-30 Hz beta activity band that appears after zolpidem washout and a 7 Hz pre-dose peak that attenuates. The median-smoothed spectrogram also reduced intermittent

signals such as the alpha at ~10Hz in the pre-zolpidem condition. However, by looking at the time traces of the EEG (**figure 4.5**) along with video recording, we observe that the 10 Hz rhythm co-occurs with eyes closed during time periods of increased activity and movement, and originates in the posterior channels, indicating that this signal behaves like alpha activity in healthy controls. Since the median doesn't pick up signals present in less than 50% of the the epochs, the difference between subject IN301 and subject IN376 alpha peaks is due to the frequency and duration of eyes-closed epochs.

Because the median-smoothed spectrogram loses intermittent signals that are relevant to analysis, it is worth applying and comparing both smoothed spectrogram approaches. Since EMG is higher during time periods where the subject is more awake, spectrograms where the EMG is removed or smoothed out might lose cortical signals that co-occur with EMG. However, this is a drawback that exists in the hand-cleaned spectrogram as well: when EMG are removed, signals that serve as markers of wakefulness are also removed. In particular, the posterior 10 Hz rhythm generated when the subject is awake with their eyes closed indicates intact anterior forebrain mesocircuit activity (Schiff, 2010). This motivates the inclusion of all segments in spectrogram calculations. However, it also demonstrates a limitation to the median statistic; namely, loss of information contained in intermittent signals. This can be mitigated by taking multiple approaches in order to capitalize on the strengths of each method.

CHAPTER 5: RECAP AND FUTURE WORK

This thesis develops a set of tools for time series analysis based on nonparametric quantile statistics, with the goal of reducing the effect of outliers on power spectral analysis, and applies them to EEG data. Although the EEG has been in use for almost a century, it continues to be a relevant tool for research and clinical use; this longevity is due both to the biophysics of EEG generation and a host of practical considerations. In particular, the analysis of EEG is informative about network dynamics, and is therefore a useful tool for assessing of patients with severe brain injury. A main limiting factor in the use of EEG, especially in this setting, is contamination of the recording from artifacts that arise from non-neural sources. These artifacts are often of large amplitude and at frequencies that overlap with signals of interest from the brain. Our approach to handling artifacts, rather than remove them, was to apply quantile statistics to the multitaper power spectral estimation method, as a more outlier-tolerant calculation, with accompanying confidence intervals and a spectral comparison. I developed the underlying statistical theory, vetted the method on artificial and real datasets, and disseminated the software to apply the method in a publicly available MATLAB package.

Impact

The standard approach to estimate EEG power spectra is based on computing the mean of the squared amplitudes of Fourier components, each computed from a separate data epochs and/or a taper. Because the standard approach relies on

combining these contributions via the mean, it is greatly affected by outliers. The robust approach leverages the robustness of non-parametric statistics to construct an alternative estimate of the power spectrum that has a greater tolerance to outliers. This has the advantage of producing a familiar easily interpretable result for those who do EEG spectral analysis, and it can easily be combined with other methods for dealing with artifacts. No additional statistical concepts beyond the median and power spectrum are needed to implement the method.

Additionally, as shown in chapter 4, comparing mean and median approaches on the same data can be used to identify intermittent signals vs. continuous signals. EEG from MCS subjects, who experience fluctuations in consciousness, exhibits temporal dynamics in tandem with behavioral changes: EEG from states of higher wakefulness and awareness is marked by a similarly normalized power spectrum, while EEG from states of reduced wakefulness shows a more abnormal EEG. Distinguish stationary processes from intermittent ones adds another dimension to the data, which may be useful in distinguishing between MCS and UWS states.

Although the method does not replace current outlier removal methods, the greater tolerance for outliers of the robust approach might allow one to bypass time-intensive data pre-processing for some datasets. This would effectively speed up analysis time, as well as reduce bias introduced by removal of all segments containing specific artifacts such as EMG.

Drawbacks

Not surprisingly, the potential advantages of the new approach is associated with several tradeoffs. As is typical of nonparametric approaches, the robustness comes at the expense of efficiency. 95% confidence intervals on the power spectrum require at least 6 data segments in order to be finite. When the number epochs is limited, the Bayes box test for spectral comparison is overly conservative, and loses true positives.

A separate consideration is that the approach is computationally intensive. Specifically, the Bayes Box test for spectral comparison has a runtime of $O(m^2n^2)$, where m and n are the number of segments in each dataset respectively. For cases of large datasets, other non-parametric statistics scale better: for example, the median permutation test, whose performances cluster with the Bayes Box.

Additionally, the method makes the assumption that features that are present only intermittently are artifact, and not reflective of underlying brain activity. As shown in chapter 4, this does not always hold. Some intermittent peaks do reflect underlying brain processes, so a robust power spectrum or spectrogram can lose relevant information (while some noise affects more than half the data segments and is not excluded by the method). This motivates the exploration of multiple analyses to ensure retention of relevant signal, but also reveals that signal that co-occurs with noise artifact is lost with standard artifact removal methods as well.

Applications

In the field of EEG

The robust spectral estimator may be particularly useful for analyzing long stretches of data, as is the case with EEG baseline recordings. These help probe network integrity without relying on behavioral or sensory EEG paradigms (Bai et al., 2017), and are therefore useful for diagnostic purposes in distinguishing between states of awareness. In (Williams et al., 2013), resting state EEG dynamics were used to show spectral shifts correlating to behavioral shifts before and after zolpidem, and in chapter 4 I showed that the robust approach can be used to reduce the effect of outliers even when all the data are used. This would be advantageous both for its potential to produce cleaner power spectra, but also to speed up analysis.

The method can be also applied to non-continuous data across trials, such as EEG recordings from auditory stimulus response and behavioral response paradigms (Sergent et al., 2016). These paradigms for assessing consciousness state probe network reactivity rather than latent connectivity. It may also provide a useful intermediary step in higher level EEG processing metrics, such as network connectivity and complexity analyses that show promise for DOC patients (Sitt et al., 2014).

Lastly, the robust spectrogram has been extended (Yan et al., 2017) to produce a visual tool for clinical seizure identification, with the novel approach of taking the median over segments, it calculates the median spectrum from each spatial

quadrant on the head and then takes the sum over all quadrants, for short time segments, to produce a spectrogram. A possible explanation for this method's effectiveness in this application is that the spatially applied median, i.e. applied over all channels within a quadrant, will pick up high-amplitude activity synchronous within the quadrant but reduce the contribution of large-amplitude asynchronous outliers, distinguishing between the two large-amplitude signal types. Using the non-robust sum over the quadrants produces a single spectrogram , which captures focal seizures in any one quadrant. The resulting tool displays one spectrogram, rather than simultaneous individual channel spectrograms as was done in previous seizure detection studies (Pensirikul et al., 2013). The ability of the tool to visualize sloped resonance bands characteristic of seizures allowed new residents in neurology to identify seizures with minimal training. A convolutional neural net has shown a similar ability to correctly discern seizures from other artifact, opening up more avenues for automated seizure detection tools. This motivates the further application of the median statistic to cases where outliers of interest behave differently from contaminant outliers.

It is worth exploring applications of the robust method for online data, such as for brain-computer interfaces (BCIs). The median can be used on small sets of data, but its robustness is mainly an advantage for larger datasets. Since BCIs need to capture spectral changes in realtime, collecting long enough windows of data to be effectively segmented and analyzed with the median spectrum might produce too big a lag in the response to the signal. However, this avenue is worth exploring as a

means of producing cleaner EEG signals for detecting responses in patients whose motor function lags behind cognitive function (Bai et al., 2017; Lazarou et al., 2018).

Non-EEG applications

At its core, the method can be applied to non-EEG data as well. The method benefits any time series spectral analysis where the signal of interest is continuous and the outliers are intermittent and biased in one direction (in EEG, outliers are predominantly large). Whenever normality of the spectral components cannot be assumed, a robust non-parametric approach is more appropriate than a parametric one. Given the interest in producing robust estimates on other time series such as geophysical and climactic (Chave et al., 1987; Kleiner and Martin, 1979), utility of the method outside of EEG is worth exploring, in particular whererever the multitaper method is applied.

Future work

Researchers and clinicians frequently use other spectral measures, such as coherence between channels, to assess connectivity and level of active input into the system, as well as directionality of information flow. Coherence has been pursued in EEG research as a marker of consciousness in DOCs (Bagnato et al., 2015; Chennu et al., 2014; Höller et al., 2013; Leon-Carrion et al., 2012; Lord and Opacka-Juffry, 2016; Nolte et al., 2004; Schiff et al., 2014). Median coherograms have been used in (Wong et al., 2011) and indicate that robust statistics have utility in these and similar analyses. One drawback to the current robust method, when applied to spectral

coherence estimates, is that it loses phase information; a robust coherence method that preserves phase information is less straightforward but would be beneficial for analysis of data with clinical significance. One such approach that would robustly estimate coherence and phase is by extending the 1-dimensional median to two dimensions: applying a non-parametric approach to estimating the orientation of a minimum volume ellipse (MVE) circumscribing an ellipsoidal cloud of points developed by Rousseeuw (Rousseeuw and Van Driessen, 1999). An MVE estimator has been implemented for MATLAB in (Hubert and Debruyne, 2010). Nevertheless, some studies such as (Lord and Opacka-Juffry, 2016) used coherence without phase information to probe network connectivity changes over a period of 6-18 months of music therapy in DOC patients. Network coherence analysis, such as that shown in the thesis by (Nauvel, 2017) has diagnostic potential for revealing changing patterns of network connectivity underlying recovery from disorders of consciousness, possibly preceding observable behavioral changes.

(Nauvel, 2017) has also shown preliminary work extending analysis on the robust spectrogram by applying dimensionality reduction and classification to the zolpidem data. Cleaner spectrograms expand opportunities to use these and other analyses with less interference from outliers.

Conclusion

With applications to neurological, psychological, sleep, and consciousness disorders, the EEG remains a vital tool for clinical and basic research. The power spectrum is a useful tool for summarizing EEG recordings in the frequency domain.

Here we introduced an approach to calculating the multitapered power spectrum that can augment existing analysis pipelines and has advantages with regard to reducing sensitivity to artifact. Since the approach reduces the influence of artifacts based on their transience, applications outside the realm of EEG are also possible. The toolkit is available in a public Github repository.

REFERENCES

Andrews, K., Murphy, L., Munday, R., and Littlewood, C. (1996). Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. BMJ *313*, 13–16.

Arvesen, J.N. (1969). Jackknifing U-Statistics. Ann Math Stat *40*, 2076–2100.

Bagnato, S., Boccagni, C., Sant'Angelo, A., Prestandrea, C., Mazzilli, R., and Galardi, G. (2015). EEG predictors of outcome in patients with disorders of consciousness admitted for intensive rehabilitation. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. *126*, 959–966.

Bai, Y., Xia, X., and Li, X. (2017). A Review of Resting-State Electroencephalography Analysis in Disorders of Consciousness. Front. Neurol. *8*, 471.

Bauer, D.F. (1972). Constructing Confidence Sets Using Rank Statistics. J. Am. Stat. Assoc. *67*, 687–690.

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. Arch. Für Psychiatr. Nervenkrankh. *87*, 527–570.

Bernardo, J.M. (2011). Bayesian Statistics. In International Encyclopedia of Statistical Science, (Springer, Berlin, Heidelberg), pp. 107–133.

Bokil, H., Purpura, K., Schoffelen, J.-M., Thomson, D., and Mitra, P. (2007). Comparing spectra and coherences for groups of unequal size. J Neurosci Methods *159*, 337–345.

Bokil, H., Andrews, P., Kulkami, J.E., Mehta, S., and Mitra, P.P. (2010). Chronux: A platform for analyzing neural signals. J. Neurosci. Methods *192*, 146–151.

Bonanni, L., Thomas, A., Tiraboschi, P., Perfetti, B., Varanese, S., and Onofrj, M. (2008). EEG comparisons in early Alzheimer's disease, dementia with Lewy bodies and Parkinson's disease with dementia patients with a 2-year follow-up. Brain *131*, 690–705.

Bordini, A.L., Luiz, T.F., Fernandes, M., Arruda, W.O., and Teive, H.A.G. (2010). Coma scales: a historical review. Arq. Neuropsiquiatr. *68*, 930–937.

Box, G.E.P. (1953). NON-NORMALITY AND TESTS ON VARIANCES. Biometrika *40*, 318–335.

Britton, J. (2016). Autoimmune epilepsy. Handb. Clin. Neurol. *133*, 219–245.

Brown, R.K.J., Bohnen, N.I., Wong, K.K., Minoshima, S., and Frey, K.A. (2014a). Brain PET in Suspected Dementia: Patterns of Altered FDG Metabolism. RadioGraphics *34*, 684–701.

Brown, R.K.J., Bohnen, N.I., Wong, K.K., Minoshima, S., and Frey, K.A. (2014b). Brain PET in Suspected Dementia: Patterns of Altered FDG Metabolism. RadioGraphics *34*, 684–701.

Campbell, I.G. (2009). EEG Recording and Analysis for Sleep Research. Curr. Protoc. Neurosci. Editor. Board Jacqueline N Crawley Al *CHAPTER*, Unit10.2.

Carboncini, M.C., Piarulli, A., Virgillito, A., Arrighi, P., Andre, P., Tomaiuolo, F., Frisoli, A., Bergamasco, M., Rossi, B., and Bonfiglio, L. (2014). A case of post-traumatic minimally conscious state reversed by midazolam: Clinical aspects and neurophysiological correlates. Restor Neurol Neurosci *32*, 767–787.

Chave, A.D., Thomson, D.J., and Ander, M.E. (1987). On the robust estimation of power spectra, coherences, and transfer functions. J. Geophys. Res. Solid Earth *92*, 633–648.

Chennu, S., Finoia, P., Kamau, E., Allanson, J., Williams, G.B., Monti, M.M., Noreika, V., Arnatkeviciute, A., Canales-Johnson, A., Olivares, F., et al. (2014). Spectral Signatures of Reorganised Brain Networks in Disorders of Consciousness. PLOS Comput. Biol. *10*, e1003887.

Childs, N.L., Mercer, W.N., and Childs, H.W. (1993). Accuracy of diagnosis of persistent vegetative state. Neurology *43*, 1465–1467.

Cruse, D., Chennu, S., Chatelle, C., Bekinschtein, T.A., Fernández-Espejo, D., Pickard, J.D., Laureys, S., and Owen, A.M. (2011). Bedside detection of awareness in the vegetative state: a cohort study. The Lancet *378*, 2088–2094.

David, H.A., and Nagaraja, H.N. (2003). Order Statistics (Wiley).

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage *34*, 1443–1449.

Dvey-Aharon, Z., Fogelson, N., Peled, A., and Intrator, N. (2015). Schizophrenia Detection and Classification by Advanced Analysis of EEG Recordings Using a Single Electrode Approach. PLOS ONE *10*, e0123033.

Fins, J.J. (2015). Rights Come to Mind: Brain Injury, Ethics, and the Struggle for Consciousness (Cambridge University Press).

Fischer, D., and Truog, R.D. (2017). The Problems With Fixating on Consciousness in Disorders of Consciousness. AJOB Neurosci. *8*, 135–140.

Forgacs, P.B., Frey, H.-P., Velazquez, A., Thompson, S., Brodie, D., Moitra, V., Rabani, L., Park, S., Agarwal, S., Falo, M.C., et al. (2017). Dynamic regimes of neocortical activity linked to corticothalamic integrity correlate with outcomes in acute anoxic brain injury after cardiac arrest. Ann Clin Transl Neurol *4*, 119–129.

Geyer, C.J. (2003). Stat 5102 Notes: Nonparametric Tests and Confidence Intervals.

Giacino, J.T., Kalmar, K., and Whyte, J. (2004). The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. Arch. Phys. Med. Rehabil. *85*, 2020–2029.

Giacino, J.T., Fins, J.J., Laureys, S., and Schiff, N.D. (2014). Disorders of consciousness after acquired brain injury: the state of the science. Nat Rev Neurol *10*, 99–114.

Gosseries, O., Schnakers, C., Ledoux, D., Vanhaudenhuyse, A., Bruno, M.-A., Demertzi, A., Noirhomme, Q., Lehembre, R., Damas, P., Goldman, S., et al. (2011). Automated EEG entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state. Funct. Neurol. *26*, 25–30.

Hampel, F.R. (1968). Contributions to the theory of robust estimation /.

Herman, S.T., Takeoka, M., Hughes, J.R., and Drislane, F.W. (2011). Electroencephalography in clinical epilepsy research. Epilepsy Behav. EB *22*, 126–133.

Höller, Y., Bergmann, J., Thomschewski, A., Kronbichler, M., Höller, P., Crone, J.S., Schmid, E.V., Butz, K., Nardone, R., and Trinka, E. (2013). Comparison of EEG-Features and Classification Methods for Motor Imagery in Patients with Disorders of Consciousness. PLOS ONE *8*, e80479.

Huber, P.J. (1963). Robust Estimation of a Location Parameter. Ann. Math. Stat. *35*, 73–101.

Hubert, M., and Debruyne, M. (2010). Minimum covariance determinant. Wiley Interdiscip. Rev. Comput. Stat. *2*, 36–43.

Humphreys, I., Wood, R.L., Phillips, C.J., and Macey, S. (2013). The costs of traumatic brain injury: a literature review. Clin. Outcomes Res. CEOR *5*, 281–287.

Jameson, L.C., and Sloan, T.B. (2006). Using EEG to monitor anesthesia drug effects during surgery. J. Clin. Monit. Comput. *20*, 445–472.

Kirschstein, T., and Köhling, R. (2009). What is the Source of the EEG? Clin. EEG Neurosci. *40*, 146–149.

Kleiner, B., and Martin, R.D. (1979). Robust Estimation of Power Spectra. J. R. Stat. Soc. Ser. B Methodol. *41*, 313–351.

Krauledat, M., Dornhege, G., Blankertz, B., and Müller, K.-R. (2007). Robustifying EEG data analysis by removing outliers. Chaos Complex. Lett. *2*.

Kreuzer, M. (2017). EEG Based Monitoring of General Anesthesia: Taking the Next Steps. Front. Comput. Neurosci. *11*.

Lazarou, I., Nikolopoulos, S., Petrantonakis, P.C., Kompatsiaris, I., and Tsolaki, M. (2018). EEG-Based Brain–Computer Interfaces for Communication and Rehabilitation of People with Motor Impairment: A Novel Approach of the 21st Century. Front. Hum. Neurosci. *12*.

Lee, P.F., Kan, D.P.X., Croarkin, P., Phang, C.K., and Doruk, D. (2018). Neurophysiological correlates of depressive symptoms in young adults: A quantitative EEG study. J Clin Neurosci *47*, 315–322.

Lehembre, R., Marie-Aurélie, B., Vanhaudenhuyse, A., Chatelle, C., Cologan, V., Leclercq, Y., Soddu, A., Macq, B., Laureys, S., and Noirhomme, Q. (2012). Resting-state EEG study of comatose patients: a connectivity and frequency analysis to find differences between vegetative and minimally conscious states. Funct. Neurol. *27*, 41–47.

Leon-Carrion, J., Leon-Dominguez, U., Pollonini, L., Wu, M.-H., Frye, R.E., Dominguez-Morales, M.R., and Zouridakis, G. (2012). Synchronization between the anterior and posterior cortex determines consciousness level in patients with traumatic brain injury (TBI). Brain Res. *1476*, 22–30.

LeVan, P., Urrestarazu, E., and Gotman, J. (2006). A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification. Clin. Neurophysiol. *117*, 912–927.

Li, Y., Ma, Z., Lu, W., and Li, Y. (2006). Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. Physiol. Meas. *27*, 425.

Lord, V., and Opacka-Juffry, J. (2016). Electroencephalography (EEG) Measures of Neural Connectivity in the Assessment of Brain Responses to Salient Auditory Stimuli in Patients with Disorders of Consciousness. Front. Psychol. *7*.

Makeig, S., Bell, A.J., Jung, T.-P., and Sejnowski, T.J. (1996). Independent component analysis of electroencephalographic data. Adv. Neural Inf. Process. Syst. 145–151.

Melman, T., and Victor, J.D. (2016). Robust power spectral estimation for EEG data. J. Neurosci. Methods *268*, 14–22.

Mitra, P.P., and Bokil, H. (2008). Observed Brain Dynamics (New York, NY: Oxford University Press).

Mitra, P.P., and Pesaran, B. (1999). Analysis of dynamic brain imaging data. Biophys. J. *76*, 691–708.

Napier, S. (2015). Perception of Value and the Minimally Conscious State. HEC Forum *27*, 265–286.

Nauvel, T. (2017). The time evolution of global brain dynamics in the human electroencephalogram: innovations in quantitative multi-variate methods and applications to disorders of consciousness. Weill Cornell Medical College.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. *115*, 2292–2307.

Nunez, M.D., Nunez, P.L., and Srinivasan, R. (2016). Electroencephalography (EEG): Neurophysics, Experimental Methods, and Signal Processing.

Pensirikul, A., Beslow, L.A., Kessler, S.K., Sanchez, S.M., Topjian, A.A., Dlugos, D.J., and Abend, N.S. (2013). Density Spectral Array for Seizure Identification in Critically Ill Children. J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc. *30*, 371–375.

Percival, D.B., and Walden, A.T. (1993). Spectral analysis for physical applications (Cambridge, UK: Cambridge University Press).

Perri, C.D., Bahri, M.A., Amico, E., Thibaut, A., Heine, L., Antonopoulos, G., Charland-Verville, V., Wannez, S., Gomez, F., Hustinx, R., et al. (2016). Neural correlates of consciousness in patients who have emerged from a minimally conscious state: a cross-sectional multimodal imaging study. Lancet Neurol. *15*, 830–842.

Prieto, G.A., Parker, R.L., Thomson, D.J., Vernon, F.L., and Graham, R.L. (2007). Reducing the bias of multitaper spectrum estimates. Geophys. J. Int. *171*, 1269–1281.

Proekt, A. (2018). Chapter Fifteen - Brief Introduction to Electroencephalography. In Methods in Enzymology, R.G. Eckenhoff, and I.J. Dmochowski, eds. (Academic Press), pp. 257–277.

Radüntz, T., Scouten, J., Hochmuth, O., and Meffert, B. (2015). EEG artifact elimination by extraction of ICA-component features using image processing algorithms. J. Neurosci. Methods *243*, 84–93.

Rousseeuw, P.J., and Leroy, A.M. (1987). Robust Regression and Outlier Detection (Wiley).

Rousseeuw, P.J., and Van Driessen, K. (1999). A fast algorithm for the Minimum Covariance Determinant Estimator. Technometrics *41*, 212–223.

Sarnthein, J., and Jeanmonod, D. (2007). High thalamocortical theta coherence in patients with Parkinson's disease. J Neurosci *27*, 124–131.

Schiff, N.D. (2010). Recovery of consciousness after brain injury: a mesocircuit hypothesis. Trends Neurosci. *33*, 1–9.

Schiff, N.D., Nauvel, T., and Victor, J.D. (2014). Large-scale brain dynamics in disorders of consciousness. Curr. Opin. Neurobiol. *25*, 7–14.

Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., Majerus, S., Moonen, G., and Laureys, S. (2009). Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. BMC Neurol *9*, 35.

Sergent, C., Faugeras, F., Rohaut, B., Perrin, F., Valente, M., Tallon-Baudry, C., Cohen, L., and Naccache, L. (2016). Multidimensional cognitive evaluation of patients with disorders of consciousness using EEG: A proof of concept study. NeuroImage Clin. *13*, 455–469.

Shi, L.-C., Duan, R.-N., and Lu, B.-L. (2013). A robust principal component analysis algorithm for EEG-based vigilance estimation. Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. *2013*, 6623–6626.

Silva, L.R., Amitai, Y., and Connors, B.W. (1991). Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. Science *251*, 432–435.

Sitt, J.D., King, J.-R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A., Cohen, L., Sigman, M., Dehaene, S., and Naccache, L. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. Brain *137*, 2258–2270.

St. Louis, E.K., Frey, L.C., Britton, J.W., Hopp, J.L., Korb, P., Koubeissi, M.Z., Lievens, W.E., and Pestana-Knight, E.M. (2016). The Normal EEG (American Epilepsy Society).

Teasdale, G., and Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. Lancet Lond. Engl. *2*, 81–84.

Thomson, D.J. (1982). Spectrum estimation and harmonic analysis. Proc. IEEE *70*, 1055–1096.

Thomson, D.J. (2007). Jackknifing multitaper spectrum estimates. Signal Process. Mag. IEEE *24*, 20–30.

Thonnard, M., Gosseries, O., Demertzi, A., Lugo, Z., Vanhaudenhuyse, A., Bruno, M.-A., Chatelle, C., Thibaut, A., Charland-Verville, V., Habbal, D., et al. (2014). Effect of zolpidem in chronic disorders of consciousness: a prospective open-label study. Funct. Neurol. *28*, 259–264.

Thornhill, S., Teasdale, G.M., Murray, G.D., McEwen, J., Roy, C.W., and Penny, K.I. (2000). Disability in young people and adults one year after head injury: prospective cohort study. BMJ *320*, 1631–1635.

Whitham, E.M., Pope, K.J., Fitzgibbon, S.P., Lewis, T., Clark, C.R., Loveless, S., Broberg, M., Wallace, A., DeLosAngeles, D., Lillie, P., et al. (2007). Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. Clin. Neurophysiol. *118*, 1877–1888.

Whitham, E.M., Lewis, T., Pope, K.J., Fitzgibbon, S.P., Clark, C.R., Loveless, S., DeLosAngeles, D., Wallace, A.K., Broberg, M., and Willoughby, J.O. (2008). Thinking activates EMG in scalp electrical recordings. Clin Neurophysiol *119*, 1166–1175.

Whyte, J., and Myers, R. (2009). Incidence of clinically significant responses to zolpidem among patients with disorders of consciousness: a preliminary placebo controlled trial. Am. J. Phys. Med. Rehabil. *88*, 410–418.

Williams, S.T., Conte, M.M., Goldfine, A.M., Noirhomme, Q., Gosseries, O., Thonnard, M., Beattie, B., Hersh, J., Katz, D.I., Victor, J.D., et al. (2013). Common resting brain dynamics indicate a possible mechanism underlying zolpidem response in severe brain injury. Elife *2*, e01157.

Wong, K.F.K., Mukamel, E.A., Salazar, A.F., Pierce, E.T., Harrell, P.G., Walsh, J.L., Sampson, A., Brown, E.N., and Purdon, P.L. (2011). Robust time-varying multivariate coherence estimation: Application to electroencephalogram recordings during general anesthesia. (IEEE), pp. 4725–4728.

Yan, P., Melman, T., Yan, S., Otgonsuren, M., and Grinspan, Z. (2017). Evaluation of a novel median power spectrogram for seizure detection by non-neurophysiologists. Seizure *50*, 109–117.

Yoav Benjamini, Y.H. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.